DOCUMENT RESUME

ED 144 715 PS 009 591

AUTHOR Love, John M.; And Others
TITLE A Process Evaluation of Project Developmental Continuity. Interim Report II, Part B: Recommendations for Measuring Program Impact.
INSTITUTION High/Scope Educational Research Foundation, Ypsilanti, Mich.
SPONS AGENCY Office of Child Development (DHEW), Washington, D.C. Early Childhood Research and Evaluation Branch.
PUB DATE 30 Jun 75
CONTRACT HEW-105-75-1114
NOTE 112p.

EDRS PRICE MF-$0.83 HC-$6.01 Plus Postage.
DESCRIPTORS Biculturalism; Bilingualism; Cognitive Development; *Early Childhood Education; Emotional Development; Health; Language Development; Measurement; *Measurement Goals; *Measurement Instruments; Nutrition; Parent Participation; Perceptual Motor Coordination; *Program Effectiveness; *Program Evaluation; Social Development; Teacher Improvement; Test Selection
IDENTIFIERS *Developmental Continuity; Project Head Start; *Social Competence

ABSTRACT

This report presents recommendations for measures to be used in assessing the impact of Project Developmental Continuity (PDC). Chapter I reviews the purpose of the impact study and presents the basic considerations guiding the selection of measures. Chapter II describes the review process that led to the final recommendations. Chapter III presents the final recommendations, along with supporting rationale, for measures to be used in the areas of: (1) social-emotional development; (2) psychomotor development, health, and nutrition; (3) cognitive and language development; (4) impact on parents and teachers; (5) bilingual/bicultural education; and (6) site-specific goals. The appendix contains one-page descriptions of the recommended measures, including information about items, administration procedures, scoring criteria, validity and reliability. (Author/JMB)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED144715

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

Project Developmental Continuity Evaluation

PS 009591

SCOPE OF INTEREST NOTICE

The ERIC Facility has assigned this document for processing to: PS TM

In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

Recommendations for Measuring Program Impact

June 30, 1975

This report was prepared for the Early Childhood Research and Evaluation Branch, Office of Child Development, Department of Health, Education, and Welfare, under Contract No. HEW-105-75-1114 (Dr. Esther Kresh, Project Officer). Views or conclusions contained herein should not be interpreted as reflecting the official opinion of the sponsoring agency.

ED144715

A PROCESS EVALUATION OF PROJECT DEVELOPMENTAL CONTINUITY

INTERIM REPORT II, PART B: RECOMMENDATIONS FOR MEASURING PROGRAM IMPACT

June 30, 1975

John M. Love
Sally Wacker
Judy Meece

Technical Assistance:

Jana Grimston
Pat Loy
Cathy Peterson
Leslie Ryan

PS 009591

High/Scope Educational Research Foundation
600 North River Street
Ypsilanti, Michigan 48197

Acknowledgments

The present volume is an updated version of the original draft report submitted to OCD in June 1975. We would like to acknowledge the advice and support of a number of persons who took the time to review all or parts of the report during the initial preparation stages and after the draft was circulated. We have benefited greatly from this multifaceted review process, but of course wish to absolve our colleagues of any responsibility for shortcomings in the final product.

Our government project officer, Esther Kresh, has been a continuing source of advice and has been of particular value in reminding us of issues pertinent to the conduct of large-scale national evaluations. Jenny Klein, who at OCD was most responsible for the initial year of PDC, was particularly helpful in clarifying program goals.

Our National Advisory Panel convened in April 1975 to review initial plans in each measurement area and provided a wide range of expert opinion. We are grateful to Frank Angel, Charles Billings, Ricardo Cornejo, Robert Egbert, Edward Gotts, and Eugene Litwak for their suggestions. Several others reviewed initial plans related to bilingual/bicultural children--Soledad Arenas at OCD, Ernesto Bernal, Luis Laosa, Esperanza Medina, and Luis Rivera. Robert Cooke and Philip Mirvis were particularly helpful in identifying issues and strategies in the measurement of teacher and staff outcomes.

Several other individuals were kind enough to review the draft report and respond both to the overall design and to specific measurement issues; we would like to thank Ed DeAvila, Edith Grotberg, Asa Hilliard, and Bernard Spolsky for taking the time to do this.

Finally, we are indebted to our many colleagues at High/Scope who participated in long and frequent discussions, answering questions and giving advice--Marilyn Adams, Bernie Banet, Terry Bond, Ann Epstein, Salvador Lopez, Sheila Mainwaring, Judy McNeil, Mary Morris, Joanna Phinney, Mel Shelly, Susan Shipstead, and Dave Weikart.

Preface

Interim Report II was submitted at the end of the first year of Project Developmental Continuity (PDC) to present our recommendations for measuring program impact. The present volume represents an updating of the June 1975 draft report. After the draft report was reviewed by OCD and various consultants, several of our recommendations were modified; these modifications have been incorporated into the present volume. The intent of this report is to provide the basic rationale of the measures to be used in PDC; it is not an exhaustive review of all possible measures.

Most of the 15 PDC programs were funded in the summer of 1974 for a year of planning during which preparations would be made for beginning operations in September 1975 in all components--education, training, parent involvement, administration, support services, services for handicapped children and programs for bilingual/bicultural children. As with all Head Start demonstration programs, Developmental Continuity is a multifaceted program that requires a complex evaluation process. In order to provide answers to many questions regarding programs and policy, the Office of Child Development designed a broad-based study that focuses both on the process of establishing and operating programs and on the impact that the programs are expected to have on children, families, Head Start and school teaching staff, administrators and on the institutions themselves.

During 1974-75, the evaluation focused on assessing the planning process. Two visits were made to each program, and a case study was prepared for each program to describe and analyze the year's events. A "national case study" provides a summary of planning activities from a national perspective and an analysis of trends across sites. The 15 program case studies and the national case study constitute Part C of this report.

The assessment of program impact, of course, can only occur after there has been an opportunity for children to experience continuity from Head Start to elementary school. Since there are frequently problems during program start-up that are not typical of a fully operating program, there will be no assessment of impact during 1975-76 (Year II); the first cohort to be tested, interviewed and observed for impact assessment will enter in the fall of 1976 (Year III). During Year II the assessment procedures will be pilot tested on a sample of Head Start children from PDC and comparison groups in most of the sites. Year I has been a year of preparation during which there have been two major concerns: locating suitable Head Start centers and elementary schools in each site to serve as "comparisons" for the PDC Head Start and school; and selecting measures that would be appropriate for assessing the impact of Project Developmental Continuity. Information on the comparison groups, along with baseline data and discussions of the analysis plan, constitutes Part A of this report.

Table of Contents

		page
Acknowledgments		iii
Preface		v
I	INTRODUCTION	1
	Purpose of the Impact Study	1
	Considerations in Selecting Measures	2
	General Guidelines	2
	Specific Criteria	4
II	PREPARATION FOR MEASUREMENT SELECTION	7
	Defining Social Competence	7
	Goals for Parents, Teachers and PDC Staff	10
	Developmental Continuity Goals	10
	Literature Review	17
III	RECOMMENDATIONS FOR IMPACT MEASUREMENT	19
	Measures of Social-Emotional Development	20
	Defining Areas of Social-Emotional Competence	20
	Feelings of individual worth accompanied by realistic self-appraisal	20
	Self-directing, feeling he/she can influence the outcome of events	21
	Learning how to learn	25
	Social problem-solving, including self-assertion	26
	Recognition of feelings in self and others	27
	Positive school attitude	28
	Alternatives to Testing	28
	Developmental Continuity observation system	29
	Summary of Recommended Measures for the Basic Battery	30
	Measures of Psychomotor Development, Health and Nutrition	32
	Defining Areas of Psychomotor Competence	32
	Perceptual performance	34
	Block Design	35
	Motor performance	35
	Visual and auditory perception	36
	Defining Areas of Health and Nutrition	36
	Recommended procedures	36
	Summary of Recommended Measures for the Basic Battery	37
	Measures of Cognitive and Language Development	38
	Defining Areas of Cognitive and Language Competence	38
	General cognitive competencies	39
	Problem-solving competencies	39
	Verbal competency	40

Table of Contents
(continued)

		page
	Measurement Selection Process.	40
	MSCA subtests.	44
	Block design	45
	Summary of Recommended Measures for the Basic Battery.	45
	Measures of Impact on Parents and Teachers.	47
	Defining Program Goals for Parents and Teachers. . . .	47
	Parent outcomes.	47
	Teacher outcomes	48
	Staff-as-a-group outcomes.	49
	Institutional change outcomes.	49
	Selecting Measures for Assessing Impact.	51
	Summary of Measurement Procedures.	52
	Bilingual/Bicultural Measures	57
	Defining Bilingual/Bicultural Measurement Areas. . . .	58
	Recommended Measures	61
	Adapting the Basic Battery for Bilingual/Bicultural Sites	63
	Measures for Site-Specific Batteries.	65
	Issues in the Development of Site-Specific Batteries .	65
	Program Goals.	66
	Reduced Importance of Site-Specific Batteries.	67
	Recommended Measures for Site-Specific Goals	68
	School readiness	68
	Productive language competence	70
	Flexibility in the application of information processing strategies	71
	Auditory discrimination.	72
	Procedures for Implementing Site-Specific Batteries. .	72
IV	SUMMARY OF RECOMMENDED MEASURES.	73
V	REFERENCES	75
	APPENDIX: INSTRUMENT REVIEWS.	83

List of Figures

Figure 1. Areas of Social Competency Defined by Anderson and Messick (1974). 8

Figure 2. Areas of Social Competency Defined for the Developmental Continuity Evaluation. 11

Figure 3. Areas of Expected Impact on Parents, Teachers, Staff, and Institutions. 12

Figure 4. "Most Important" Goals for Children as Rated by 11 PDC Programs 14

Figure 5. "Most Important" Goals for Staff and Parents as Rated by 11 PDC Programs 16

Figure 6. Recommended Categories for the PDC Classroom Observation System 31

Figure 7. A Model for Viewing PDC Impact 53

Figure 8. Types of Assessment Procedures for Measuring Impact on Parents, Teachers, Staff, and Institutional Change 54

Figure 9. Instruments for Assessing Perceptions and Attitudes. 55

List of Tables

Table 1. Ratings of Measures of Social-Emotional Development on Criteria for Measures Selection. 22

Table 2. Ratings of Measures of Psychomotor Development on Criteria for Measures Selection 33

Table 3. Ratings of Measures of Cognitive and Language Development on Criteria for Measures Selection . 41

Table 4. Summary of Measures for Cognitive and Language Basic Battery. 43

Table 5. Ratings of Measures for Spanish-Speaking Children on Criteria for Measures Selection 62

I

INTRODUCTION

This volume presents the recommendations for measures to be used for assessing the impact of Project Developmental Continuity (PDC). In the introduction (Chapter I) the purpose of the impact study is reviewed and the basic considerations guiding the selection of measures are presented. Chapter II describes the review process that led to the final recommendations, which are presented in Chapter III with supporting rationale. The appendix contains one-page summaries of the characteristics of the measures recommended with descriptions of the items, administration procedures and scoring criteria.

Purpose of the Impact Study

The work statement for the Project Developmental Continuity evaluation (RFP-4-75-HEW-OS) specified two primary objectives for the impact study:

- To assess the impact of Preschool-School Linkages and Early Childhood Schools on the development of social competence in children four to eight years of age; and

- To assess the impact of the two program approaches (ECS and PSL) on the preschools, schools, and other community institutions involved in the program.

The impact of Developmental Continuity must obviously be broadly conceptualized, not in terms of children's "intelligence," or even solely in terms of the children--but in terms of all areas in which the program might be expected to have an impact. For convenience, the areas of expected impact can be grouped into four categories:

- The institutions and their relationships
- PDC staff, teachers and administrators
- Parents
- Children's social competence

A variety of measures are required to complete an assessment in each of these areas. The impact on the participating institutions and any changes in their relationships will be assessed

primarily through structured and semi-structured interviews with key people connected with the program, by asking participants in key roles to rate factors such as the influence of the participants, and through examination of records. Much of this is being accomplished in the program case studies (see Part C of this report) but some additional measures, not yet part of the case study procedures, are reviewed here.

The impact on PDC staff, teachers and administrators and on parents will be assessed using interviews, questionnaires, and rating scales and by examining records such as minutes of committee meetings. A wide range of effects is expected for these participants (participation in decision-making, problem-solving, attitude change, etc.), and these are also discussed below.

The impact of PDC on children's social competence is probably the most complex of the areas of expected impact and the area that appears to receive the greatest attention in this report. This is due to the fact that social competence is itself a multifaceted concept, that its conceptualization and operationalization are subject to considerable controversy, and that a wide selection of assessment procedures is available for review. The four dimensions of social competence considered here (social-emotional, psychomotor, cognitive and language, and health and nutrition) can be assessed by interviews, rating scales, structured tests, classroom observation and by examining program records in an attempt to obtain as full a picture as possible of the developing "competence" of children who experience Developmental Continuity.

Considerations in Selecting Measures

Six general "guidelines" have been followed in making decisions on the measures for this evaluation. In addition, five specific criteria have been applied in assessing the suitability of the measures. In this discussion a distinction is made between the "basic" measurement battery--a set of measures that will be used in all PDC sites to assess objectives that are characteristic of PDC as a national program--and site-specific batteries relevant to the unique objectives of individual programs and suitable for the cultural setting of each site.

General Guidelines

The first guideline is that no new measures will be scheduled for development as part of the basic battery, other than questionnaires and interview protocols specific to this study. It is possible to distinguish studies which are intended to develop tests from studies intended to produce

information for substantive decisions (Cronbach, Gleser, Nanda, Rajaratnam, 1972). The two kinds of studies must be designed differently for maximum effectiveness and efficiency, and in general yield less than optimal results when combined into a single study. This would argue against selecting new instruments, or promising but relatively undeveloped ones which would need major instrument development, as an integral part of the evaluation. In addition, adequate test development is a prohibitively expensive undertaking that is beyond the available resources of this evaluation.

Unfortunately, as Chapter III will indicate, no measure that is already fully developed has been found that meets all the specific selection criteria listed below. Further, many evaluations (e.g., the national Follow Through evaluation) have been criticized precisely because their use of existing standardized tests did not yield the kind of information useful for policy decisions (Rivlin and Timpane, 1975). Therefore, this first guideline was modified as follows: subtests or sets of items of developed measures that best meet the criteria are recommended, and modifications of less well-developed measures are being recommended for pilot testing during Year II.

Thus, a second guideline comes into play: some instrument development can be anticipated for each of the measures in the proposed battery. Even though major test development is not feasible, nearly all of the measures considered for inclusion in the evaluation could profit from further development of a relatively minor and straightforward nature. For example, this may include something as simple as augmenting already existing normative data, or as complex as item revision and factor rearrangement. If absolutely necessary, say if one of the proposed measures does not work out as expected based on analysis of the Year II fall data, it would still be possible to substitute an entirely new measure for the spring data collection of Year II. It is hoped, however, that the process will be one of continuous refinement, rather than a series of major changes.

Third, instruments must be evaluated for the basic battery in terms of the overall objectives of Project Developmental Continuity, and must be evaluated for the site-specific batteries in terms of site objectives and ethnic and bilingual/bicultural considerations. As part of the measurement selection process a questionnaire was developed to obtain information from each site on the program goals or objectives considered especially important at that site. The responses to this questionnaire and the decisions based on these responses are discussed in Chapter II.

Fourth, out of consideration for participants in the Developmental Continuity program, testing time for any individual should be limited to a reasonable length. This suggests that the different measures used should be the shortest possible versions consistent with sound measurement practices. Moreover, it is not feasible to try to "zero in" on a particular trait in this evaluation by administering several different but overlapping measures of the trait in order to utilize, for example, the multitrait-multimethod matrix technique. Some measures will be unobtrusive as far as the child is concerned, and are therefore limited only by adequacy of records or amount of parent, teacher, or administrator time they would take.

Fifth, the measurement battery should be as simple and parsimonious as possible to insure the accuracy of data collected under different field conditions. Many elaborate but otherwise good tests must be removed from consideration for this reason.

Sixth, single items should be individually interpretable, to the extent possible, regardless of any score or factor structure they were intended to be embedded in. One can never be sure that the reliability, validity, factor analysis outcomes, test ceilings and floors, etc., will hold up for the population in this project. This suggests that the safety factor of being able to interpret individual items as straightforward criteria might ultimately prove to be a very useful feature. The more simple and straightforward each proposed test is at the item level, the greater the probability of getting some interpretable results at the end of the project even under the worst of circumstances.

Specific Criteria

In reviewing candidates for inclusion in the proposed battery, five criteria were used to assess the suitability of measures. In order of importance, the measures must:

- appear to measure stated national or local objectives;
- be appropriate to the examinees' age, ability level, and bilingual/bicultural status;
- be practical to administer;
- have been used successfully in other major evaluations (for the basic battery);
- demonstrate good psychometric characteristics.

Each of these criteria is discussed in turn below.

First, measures have been carefully examined for item content to see if they appear to measure traits related to the goals of the national or local project. The inspection of content has been used as the primary evidence for validity because of a general lack of quantitative validity data for most of the measures under consideration.

Second, measures which are relevant to the goal areas, have been reviewed in terms of their suitability for the Project Developmental Continuity population. This means looking for the floor and ceiling levels to ensure an adequate 4- to 8-year age range, an examination of procedures and format for ease of understanding by the children, and an examination of suitability of use with bilingual/bicultural children.

Third, measures have been examined to determine if they are practical to administer. The most important criterion here is the overall time required by the test, but another very important one is the general suitability for administration by paraprofessional testers hired from the community. This criterion has eliminated many measurement techniques that require formal credentials, including the better known individual IQ tests, projective tests, and depth interviews. Ease and clarity of scoring is another quality sought in the measures.

Fourth, major past uses of the measures have been taken into consideration, with preference given (for use in the basic battery) to those used in other national evaluations. The search for measures began with those used in the national Planned Variation Head Start evaluation conducted by Stanford Research Institute and with those used in the evaluation of Home Start conducted by High/Scope Educational Research Foundation. The purpose of this criterion is to establish, insofar as possible, comparability of data with those two major evaluations of preschool-aged children. Since a great deal of comparability is not likely given the years of test development work since the inception of those projects, the recommendations of the Rand Corporation study (Raizen and Bobrow, 1974a) have been considered and the measures currently being considered by the Stanford Research Institute for the National Day Care Cost Effects Study have been reviewed.

Fifth, measures that seem satisfactory in regard to the four criteria mentioned above have been examined for reliability, excellence of norms, method of construction, and suggestive results in other studies--in short, in all their important technical characteristics. It might seem peculiar to examine such important features last, but if measures failed to meet some of the more pragmatic criteria above there was no need to consider them further--good psychometric qualities notwithstanding.

II

PREPARATION FOR MEASUREMENT SELECTION

The process of reviewing measures to be used for assessing Project Developmental Continuity impact varied according to the goal areas to be assessed: institutional change, effect on teachers and other staff, impact on parents, and impact on children's social competence. The program case studies have, since the mid-year site visits, been collecting information that is relevant to questions of institutional change and teacher and staff impact. Since those issues are dealt with-in the case study volume (Part C) they will not receive as thorough a review here as the areas of impact on parents and children. In Chapter III, however, measures recommended in addition to the case study procedure will be described (see section on "Measuring Impact on Parents and Teachers"). Thus, Chapter II focuses on attempts to delineate the critical dimensions of "social competence" in children and on the process underlying the review and selection of measures for assessing impact on children, parents, teachers and staff.

Defining Social Competence

In January 1973 a panel of experts met at the Educational Testing Service to attempt to define the meaning of social competency in young children (Anderson and Messick, 1974). The deliberations of the panel culminated in a list of 29 competencies (see Figure 1). The nature of this list is important to the considerations for measuring social competence in Developmental Continuity--the list represents child behaviors that span a broad range. The rationale for this is in part attributable to the views of Edward Zigler, who in 1972 described Head Start as hoping:

> ...to bring about greater social competence in disadvantaged children. By social competence is meant an individual's everyday effectiveness in dealing with his environment...his ability to master appropriate formal concepts, to perform well in school, to stay out of trouble with the law, and to relate well to adults and other children (quoted by Anderson and Messick, 1974, p. 283).

Figure 1

Areas of Social Competency Defined by Anderson and Messick (1974)

Differentiated self-concept and consolidation of identity
Concept of self as an initiating and controlling agent
Habits of personal maintenance and care
Realistic appraisal of self, accompanied by feelings of personal worth
Differentiation of feelings and appreciation of their manifestations and implication
Sensitivity and understanding in social relationships
Positive and affectionate personal relationships
Role perception and appreciation
Appropriate regulation of antisocial behavior
Morality and prosocial tendencies
Curiosity and exploratory behavior
Control of attention
Perceptual skills
Fine motor dexterity
Gross motor skills
Perceptual-motor skills
Language skills
Categorizing skills
Memory skills
Critical thinking skills
Creative thinking skills
Problem-solving skills
Flexibility in the application of information-processing strategies
Quantitative and relational concepts, understandings, and skills
General knowledge
Competence motivation
Facility in the use of resources for learning and problem-solving
Some positive attitudes toward learning and school experiences
Enjoyment of humor, play, and fantasy

In 1973 the Rand Corporation was asked to design a national evaluation of Head Start, and this task included recommending measures to assess social competence (Raizen and Bobrow, 1974a). The result was a large set of dependent variables covering a narrow range of socially competent behaviors (for which there were not always adequate measurement procedures). Even with the large number of measures recommended by Rand, only certain aspects of social competence were included. The report points out that the definition of social competence must be a function of the specific roles that are required of the child and must be measured in relation to the context in which those roles are manifested. By arguing that Head Start is a "preface" to the role of pupil and that Head Start-eligible children are probably less prepared for the role of pupil as defined by the dominant culture, the Rand report restricted the definition of social competence to "effectiveness in the role of pupil" (Raizen and Bobrow, 1974a, pp. 17-18).

Although the role of pupil is clearly a concern of the architects of Developmental Continuity, restricting social competence to competence in that role would seem to be too narrow when compared to the more inclusive list of competencies described by Anderson and Messick, and when compared to the Office of Child Development's concern with the child's "everyday effectiveness in dealing with his environment and responsibilities in school and life." Competency within the school may entail quite different skills from competency in the out-of-school environment, and it seems that successful accommodation by the child to school is not sufficient to insure the low income child the option of successful participation in his/her home and out-of-school environment. The Head Start-elementary school cooperation required for Developmental Continuity also provides for the possibility that the schools might have to change in order to enhance a child's social competence. Thus, emphasizing "effectiveness in the role of pupil" might not allow for a change in the definition of that role.

With these considerations in mind, a definition of social competence was adopted for this evaluation based on the areas described by Anderson and Messick (1974). Although their definition is relevant to everyday effectiveness in the majority culture of the United States, it attempts to be fundamental enough to avoid criticisms of cultural bias by identifying behaviors which are more universally functional. That is, behaviors and attitudes have been identified which are appropriate for the child's effective functioning both at school and at home and which are also cornerstones for adult mental health. Although there is as yet no developmental theory of affective growth comparable to that of cognitive growth in which to anchor these attitudes and behaviors, qualities have been

identified that would be an asset in most social settings and that would probably enhance cognitive performance. Nevertheless, it is recognized that there are certain cultural values that could be part of a child's home life that may come in conflict with some of the behaviors valued by Developmental Continuity. For example, a child who is taught unquestioning respect for authority at home may have some difficulty when encouraged to be more "self-directing" by teachers. This conflict may not be resolved, but should be acknowledged.

The categories of social competence have also been influenced by the work of White and Watts (1973) at Harvard and the categories of behavior used in the observational instruments developed by Ogilvie and Shapiro (1972) and Martha Bronson (1975). The result of reviewing these different attempts to define social competence was the identification of 23 areas classified under five dimensions: six broad areas of social-emotional competence, five of psychomotor competence, five of language, six of cognitive, and one for health and nutrition. These areas are listed in Figure 2.

Goals for Parents, Teachers and PDC Staff

Of particular concern in this demonstration project is the effect that Developmental Continuity might have on parents, on teachers, and on other Head Start and school personnel. Developmental Continuity expects teachers to become more involved in a cooperative process of educational planning and to become more aware of the needs of children at all levels from Head Start through grade three. Parents are expected to become more involved in the educational system, in terms of participation in the educational process and in decision-making roles. Teacher and parent attitudes might also be expected to change. Opinions of OCD program staff and of several consultants were incorporated with information from previous evaluation studies to develop a list of important goals for parents and teachers. These are presented in Figure 3 and are discussed in the appropriate sections of Chapter III.

Developmental Continuity Goals

Although the definition of measures for this evaluation has benefited from the previous work just described, it is not likely that the definition of social competence developed for the Developmental Continuity evaluation, or the way in which it is operationalized, will be <u>generally</u> acceptable beyond the PDC context. This is partly because, in addition to conceptual considerations, the particular goals and concerns of Project Developmental Continuity on both the national and local levels,

Figure 2

Areas of Social Competency Defined for the Developmental Continuity Evaluation

Social-Emotional Development

- Realistic self-appraisal, accompanied by feelings of individual worth
- Self-directing; feels he/she can influence the outcome of events
- Learning how to learn
- Social problem-solving
- Recognition of feelings in self and others; sensitivity to and understanding of others
- Positive school attitude

Psychomotor Development

- Fine motor dexterity
- Gross motor skills
- Perceptual skills
- Perceptual-motor skills
- Auditory skills

Language Development

- General language use
- Aural comprehension
- Descriptive competence
- Functional competence
- Productive competence

Cognitive Skills

- Categorizing
- Memory
- Problem-solving
- Flexibility in the application of information processing strategies
- Quantitative and relational understanding skills
- School readiness

Health and Nutrition

- Child develops better eating habits, is free from disease and attains a level of positive health

Figure 3

Areas of Expected Impact on
Parents, Teachers, Staff and Institutions

Parent Outcomes

- Participates in school and classroom activities
- Demonstrates understanding of a child's developmental educational process
- Provides input into school decision-making and problem-solving activities
- Increases personal development through participation as a decision-maker and problem-solver

Teacher Outcomes

- Provides instruction matching child's developmental learning level
- Creates classroom environment conducive to development of social competency
- Acquires knowledge of resources available to meet needs of PDC child and families
- Provides activities for bilingual/bicultural children

Staff-as-a-Group Outcomes

- Staff members interact with members of differing grade levels
- Staff members plan and develop jointly educational and program goals for children, parents and themselves
- Staff members commonly gain competencies in group problem-solving and decision-making techniques

Institutional Change Outcomes

- Consistent staff and parent interaction
- United effort by teachers, staff and parents in determining educational goals consistent across grade levels
- Integration of Head Start and Elementary school philosophies and services

have been considered in developing the measurement strategy. To obtain more detailed information on program goals than was available in the program Guidelines, a "PDC Program Goals Questionnaire" was developed. The questionnaire contained statements describing the 25 child social competencies listed in Figure 2, 18 teacher or PDC staff goals, and 10 goals for parents. Two types of responses were elicited. First, staff were asked to rate each of the individual goals on the following scale: "will emphasize," "important but not possible," or "will not emphasize--not important." Second, they were asked to list the five "most important" goals for children and the five "most important" for teachers and parents. The questionnaire was explained to each of the PDC project coordinators at a national meeting in May 1975, at which time they were asked to work with their staff to complete the questionnaire in a way that would represent the opinions of the local program. In addition, national OCD staff were asked to complete the questionnaire in terms of the national PDC perspective on goals and priorities.

Eleven sites completed and returned the questionnaire; the national office did not. Thus, the priorities discussed here represent local program goals, although national OCD input was obtained initially and influenced the statements that were included in the questionnaire. The purpose of the analysis of responses to the goals questionnaire is twofold: first, where there is considerable agreement among programs, this provides evidence that those goals should be assessed in the basic measurement battery; second, where there is disagreement, this suggests that site-specific measures should be introduced so that the evaluation can be relevant to local concerns. The first analysis of responses was in terms of the goals rated as most important. Figure 4 lists the goals selected as most important for children according to the frequency with which sites selected them.

The sites that responded ranked five of the six social-emotional aspects of social competence among the top eight goals that their program intended to emphasize. None of the five psychomotor goals was included among the top eight. Of the five language goals, only general language use was emphasized, and it ranked first overall with 8 of the 11 sites listing it among their five most important goals. Of the six cognitive goals, problem-solving skills was considered to be the most important and it was ranked second overall among the child goals (listed by 6 sites). Three of the other cognitive goals were not considered "most important" by any site, and the other two were important in only one or two sites. Health and nutrition and self-directing were tied for third in the rankings. These were followed by the social-emotional goals of learning how to learn, realistic self-appraisal, recognition of feelings, and social problem-solving.

Figure 4

"Most Important" Goals for Children as Rated by 11 PDC Programs[1]

Child Goals Listed as "Most Important" by Five or More Sites:

- General language use
- Problem-solving skills
- Self-directing; feels he/she can influence the outcome of events
- Health and nutrition

Child Goals Listed as "Most Important" by Two to Four Sites:

- Learning how to learn
- Realistic self-appraisal, accompanied by feelings of individual worth
- Recognition of feelings in self and others; sensitivity to and understanding of others
- Social problem-solving
- Positive school attitude
- Gross motor skills
- Perceptual motor skills

---

[1]Listed in order of the number of sites selecting the goal, with the goal selected by the greatest number listed first.

Of the 28 teacher, staff and parent goals, the staff goal ranked most important by the greatest number of sites (6 out of 11 sites) was the goal that staff be better able to individualize their approach to children (see Figure 5). Parent input in decision-making regarding the total school program was listed as the most important goal to parents by 6 sites and parent input into decisions regarding the program for their child was listed by 5 sites.

Another perspective on program goal priorities was obtained from the part of the questionnaire that asked program staff to rate each individual goal on the three-point scale described above. A number of programs saw various social-emotional goals as impossible (even though they might be important goals), particularly the goal of children becoming more self-directing. Another area frequently seen as impossible was the cognitive area where all but one of the goals was seen as impossible by one site or another. Given the emphasis on health and nutrition in Head Start and in PDC, it was surprising to find two programs citing health and nutrition goals as not possible to achieve. On the other hand, all sites saw the psychomotor goals as possible to achieve, whereas it might be expected that educational programs would have little influence on psychomotor development. Only two sites saw any staff goals as impossible, but one site did list seven of the 18 staff goals as impossible to achieve. The goals for parents were almost unanimously seen as possible accomplishments.

Another outcome of the goals questionnaire was an indication of the goals that would not be emphasized or would be considered as not important by the programs. Almost half (5) of the sites responding to the questionnaire said that none of the goals was "not emphasized." The sites that did consider some of the goals to be unimportant were most likely to feel that way about the social-emotional, language and cognitive areas. The least important social-emotional goals, judging from these responses, seemed to be competence motivation, learning how to learn, and recognition of feelings in self and others. The responses to the language area were largely non-bilingual programs responding that goals regarding bilingualism were not important. In the cognitive area, three sites considered four or more of the cognitive goals to be unimportant; these tended to be the areas of categorizing, memory, problem-solving, and flexibility in the application of information processing strategies.

Figure 5

"Most Important" Goals for Staff and Parents as Rated by 11 PDC Programs[1]

Staff and Parent Goals Listed as "Most Important" by Five or More Sites:

- Staff is better able to provide differing experiences for children, matched to the children's needs; staff utilizes resources appropriate for individual children
- Parents have more input into decisions about the total school program for all children
- Parents have more input into decisions about the school program for their child

Staff and Parent Goals Listed as "Most Important" by Two to Four Sites:

- Staff has greater awareness of child development in terms of emotional needs and social competencies from age 4 to 8 years
- Parents feel more comfortable in interacting with school staff
- Parents participate in more school-related activities
- Staff is more aware of total PDC program for children aged 4 to 8 years, rather than just the area of their own responsibilities
- Staff is aware of and uses methods of positive social reinforcement for children
- Staff members plan jointly for individual children and for groups of children, developing commonly held and understood values
- Parents participate in more school-related activities
- Staff interacts more with parents of children in the classroom, at meetings, in home visits

---

[1]Listed in order of the number of sites selecting the goal, with the goal selected by the greatest number listed first.

Literature Review

The problem of selecting the most appropriate instruments for measuring the aspects of social competence designated most important was addressed within the general constraints of evaluation research discussed in Chapter I. Measurement considerations that are specific to the methodological or conceptual issues in a particular area will be discussed within each category of social competence in Chapter III.

Once the important aspects of social competence were delineated, a thorough literature search was initiated to find appropriate measures. This included the following sources:

- Research reports on Planned Variation Head Start (e.g., Smith, 1973)
- Reports of the Home Start evaluation study (e.g., Deloria, et al., 1974)
- Evaluations of the national Follow Through project (e.g., Sorenson and Madow, 1971; Emerick, Sorenson and Stearns, 1972; Cline, 1974)
- The Rand Corporation report on social competence (Raizen and Bobrow, 1974a)
- Instrument reviews (e.g., Butler, et al., 1971; Walker, 1973; Johnson and Bommarito, 1971; Robinson and Shaver, 1974; Buros, 1965, 1972; Hoepfner, Stern and Nummedal, 1971)
- ETS Head Start longitudinal study (e.g., Anderson, et al., 1968)

Numerous other articles, reports and papers were reviewed; many of these are referred to in the discussions of specific measurement areas in Chapter III and are included in the references.

III

RECOMMENDATIONS FOR IMPACT MEASUREMENT

This chapter presents the background, rationale and recommendations for the measures to be used in assessing the impact of Project Developmental Continuity. The discussion is organized by the broad program goal areas: social-emotional development; psychomotor, health and nutrition; cognitive and language development; impact on teachers and parents; and bilingual/bicultural impact goals. For each of these areas the important constructs are described, the potential measures are discussed, and the reasons for selecting and rejecting measures are reviewed.

A description of each recommended measure, with information on its reliability, validity, practicality, etc., is included in the appendix.

Measures of Social-Emotional Development

Defining Areas of Social-Emotional Competence

As indicated in Chapter II, the comprehensive paper by Anderson and Messick provided the foundation for identifying the most important aspects of child social-emotional competence. Also influential were the Rand report and the findings of Burton White, Bernice Shapiro, and Martha Bronson at Harvard. After reviewing these sources, six broad categories of social-emotional competence of preschoolers were derived:

- Feelings of individual worth accompanied by realistic self-appraisal;
- Self-directing, feeling he/she can influence the outcome of events;
- Learning how to learn, competence motivation;
- Recognition of feelings in self and others;
- Social problem-solving, including self-assertion;
- Positive school attitude.

A discussion of each of these areas, and possible measures for them, follows. Table 1 provides a summary of how each measure was judged according to the selection criteria:

Feelings of individual worth accompanied by realistic self-appraisal. This area, commonly summarized by the term "self-concept," includes the avoidance of extremely negative self-depreciation even if the child is at a rather low level of proficiency, as well as an awareness and appreciation of his/her own cultural background. The importance of positive self-concept and realistic self-appraisal has long been recognized by child development specialists, Head Start staff and OCD. However, the difficulty of obtaining a valid indicator of self-concept, especially for young children, has become increasingly apparent.

The fundamental difficulty with the construct of self-concept is the multiplicity of definitions. Many critics (Coller, 1971; Crowne and Stephens, 1961) have concluded that what are called self-concept measures actually assess many different constructs. No agreed upon operational definition of self-concept exists, largely because there is no adequate theory within which it can be based. As a result there are a great variety of instruments and techniques for narrowly defined aspects of self-concept, so that "it is likely to be defined as 'that which a self-concept test measures'" (Walker, 1973). This, of course, makes the selection of any one instrument as a measure of program effectiveness especially problematic.

Several measures either recommended or used in large-scale evaluations were examined and found wanting (see Table 1):

- Children's Self-Social Constructs Test: Preschool Form, Self-Esteem Subtest (Long and Henderson, 1970);
- Tennessee Self-Concept Scale (Fitts, 1964);
- Piers-Harris Self-Concept Scale (Piers and Harris, 1969);
- Brown IDS Self-Concept Reference Test (Brown, 1966);
- Faces Scale (in Walker, 1973).

The first was recommended for use in the national battery by Rand in their final report; it was rejected for use in PDC on the basis of questionable validity. The child is asked to choose a circle from a column of five circles; it is unclear that choosing a circle closer to the top represents "higher" self-esteem for a child. The second two were recommended by Rand in their preliminary report and dropped in their final report, presumably because they are designed for an older population and would require too much development for preschoolers.

The Brown Self-Concept Test has been widely used (Boger and Knight, 1969; Walker, Bane, and Bryk, 1973; Emrick, et al., 1972; Shipman and Gilbert, 1972). Boger and Knight questioned the extent to which their data were confounded by the effects of social desirability, and Walker, et al., stated that "three large-scale studies found that the scores were negatively skewed and showed a ceiling effect, indicating that young children have a strong tendency to select positive, 'socially acceptable' attributes." It was also felt that this was more of a cognitive or language test for younger children (who have trouble with longer items and double negatives) and that it was culturally biased (Walker, et al., 1973). Therefore its appropriateness for PDC is doubtful.

The Faces Scale is designed to measure self-concept with regard to school and was used in two pilot Follow Through studies (Emerick, et al., 1972). Correlations with the Brown and other social-emotional measures were low, but more importantly, reliability indicators were low.

The inadequacies of these available measures and a reluctance to engage in extensive pilot work on a construct which is so difficult to define made the choice of structured observations on specifiable behaviors appear to be the most valid measure of self-concept. Thus, it was decided to design an observation procedure for this evaluation. The initial steps in developing the PDC Classroom Observation System are described below.

Self-directing, feeling he/she can influence the outcome of events (locus of control). This involves the child's setting his/her own goals, taking some responsibility for skill acquisition and self-care, expecting that his/her behavior could change the probability that reinforcement would occur (that is, feeling responsible for the positive or negative responses of others), and

Table 1

Ratings of Measures of Social-Emotional Development on Criteria for Measures Selection

Measure	Measurement Selection Criteria: Practical Considerations					Psychometric Qualities				Relevance to PDC				Past Use
	Available for use by fall 1975	Appropriate for trained paraprofessionals	Test format appropriate for PDC age group	Scoring procedures appropriate for data processing	Reasonable testing time for young children	Adequate construct and/or predictive validity	Adequate test stability and internal consistency	Culture and/or SES fair	Representativeness of standardization sample	Spans age range of PDC population (4-8)	Spanish adaptation available	Relevant to PDC social-emotional goals	Likely to demonstrate program effects	Used in previous national evaluation or large-scale studies
Children's Self-Social Constructs Test		X	X	X	X		X	X	X			X		
Tennessee Self-Concept Scale	X	X		X		X	X	X				X		
Piers-Harris Children's Self-Concept Scale	X	X		X		X	X	X				X		
Brown Self-Concept Test	X	X		X	X				X			X		X
Faces Scale	X	X	X	X	X			X	X			X		X
Intellectual Achievement Responsibility Questionnaire	X	X		X		X	X	X				X		
Modified IARQ		X		X	X			X				X		
Nowicki-Strickland Locus of Control Scale		X		X	X	X	X	X				X		
Stephens-Delys Reinforcement Contingency Interview	X	X	X	X	X	X	X	X	X	X	X	X	X	X

continued:

Table 1, (continued)
Ratings of Measures of Social-Emotional Development on
Criteria for Measures Selection

Measure	Practical Considerations					Psychometric Qualities				Relevance to PDC				Past Use
	Available for use by fall '975	Appropriate for trained paraprofessionals	Test format appropriate for PDC age group	Scoring procedures appropriate for data processing	Reasonable testing time for young children	Adequate construct and/or predictive validity	Adequate test stability and internal consistency	Culture and/or SES fair	Representativeness of standardization sample	Spans age range of PDC population (4-8)	Spanish adaptation available	Relevant to PDC social-emotional goals	Likely to demonstrate program effects	Used in previous national evaluation or large-scale studies
ETS Locus of Control Scale	X	X		X	X	X			X			X		X
Preschool Interpersonal Problem-Solving (PIPS)	X	X	X	X	X	X	X					X	X	
Pupil Observation Checklist	X	X	NA	X	NA	X	X		X	X		X	X	X
CIRCUS Behavior Inventory	X	X	NA	X	NA	X	X					X		
Stanford-Binet Face Sheet	X	X	NA	X	NA	X	X		X	X		X	X	

feeling he/she can direct his/her own behavior, avoiding a feeling of powerlessness. Locus of control has become increasingly popular with researchers, in part because of its wide applicability to a variety of phenomena and its social relevance (Robinson and Shaver, 1974). The relationship of internal locus of control to academic achievement and other valued educational outcomes (Coleman et al., 1966; Nowicki and Roundtree, 1971; Rotter, 1966; and Chance, 1972) makes it of particular interest to educational evaluators. Moreover, a child's ability to become self-directing seems intrinsically worthwhile to many people and was ranked highly among individual site's responses to the PDC Goals Questionnaire.

A number of measures were considered for use or for modification:

- Intellectual Achievement Responsibility Questionnaire (Crandall, Katovsky and Crandall, 1965; modified by Ringelheim, Bialer and Morissey, 1970)
- ETS Locus of Control Scale (Hess, Shipman, Brophy and Bear, 1969)
- Nowicki-Strickland Locus of Control Scale (Nowicki and Strickland, 1973)
- Stephens-Delys Reinforcement Contingency Interview (Stephens and Delys, 1973a)

The IARQ is a well-known scale designed to tap children's beliefs in their control over and responsibility for intellectual academic success and failures; it is suitable for grades 3 through 12. The modified IARQ of Ringelheim, et al. is a shortened version of the IARQ developed specifically for use with the mentally retarded by using simpler language. Both measures employ a forced-choice format which has been a source of dissatisfaction for some researchers (Robinson and Shaver, 1974) because of the difficulty children have in responding to it. Crandall pointed out that children below grade 3 could not keep the item and the two alternatives in mind long enough to make a meaningful response (Crandall, et al., 1965). The ETS Locus of Control Scale, although employing cartoon-style drawings to aid the younger child (it is aimed at 5-1/2 to 7-1/2 year olds), suffers from the same forced-choice format, and therefore was deemed unsuitable for PDC.

The Nowicki-Strickland Locus of Control Scale is a paper-pencil "yes-no" type test designed for grades 3 through 12. Although it was judged by Robinson and Shaver (1974) to be the "best measure of locus of control as generalized expectancy presently available for use with children," it was found to be more difficult for the lower grades and therefore inappropriate for preschool children.

The Stephens-Delys Reinforcement Contingency Interview was selected because it was developed specifically for preschool children to tap a child's expectancy that his/her own behavior

would change the probability that reinforcement might occur; that is, if he/she changes his/her behavior, will his/her teachers', parents', or friends' behavior change as a result. It uses a free-response structured interview format allowing the child to use his/her own language. Forty items have been used, but a shortened form of 20 items was selected for PDC purposes. Each item poses a reinforcing-type event, e.g., "What makes mothers smile?" Responses are considered "internal" if the child answers, "When I...", and "external" when answered, "When Daddy (or someone other than self)...." The instrument has been used quite extensively with Head Start and Follow Through children (Stephens and Delys, 1973b). Because of the appropriate format of this measure for preschool children, its previous use with Head Start and ethnically diverse children, and its relatively adequate psychometric characteristics, the conclusion of Robinson and Shaver (1974) that this is the "most viable" measure of locus of control seems warranted and its use in this evaluation justified.

Learning how to learn. This third category of social-emotional competence involves the child's maintaining curiosity about his/her environment and developing an interest in school-related endeavors. It also includes learning how to follow directions, how to attend to relevant areas for an appropriate length of time and becoming aware of what he/she doesn't know, as well as developing skills in using objects and people to obtain the answer. The child should be able to ask questions of staff, of parents, and of other children, and remain involved in a particular activity until some conclusion is reached. Competence motivation, or the desire for mastery of academic skills in the absence of external pressure or rewards and the desire for competence in non-academic situations, falls within this category also.

This obviously is a broadly conceived category which owes a great deal to Bronson's nonsocial "executive skill" concept, defined as "skill in choosing and coping with tasks," but it also encompasses the dynamic aspect of motivation for competence and curiosity. This construct shares the same difficulties of assessment applicable to self-concept. The lack of relevant theory against which to judge the construct and the dearth of appropriate instruments to assess learning how to learn in preschool children supports again the conclusion that systematic observation techniques are the most meaningful measures of behaviors relevant to this construct. The Rand report recommended that "structured classroom observations directed specifically at child-task interactions" be employed to obtain a valid indicator of learning how to learn. This is also recommended for the PDC evaluation, if it is feasible to include categories representing this dimension in the observation instrument which is being developed.

Learning how to learn strategies might also be evidenced in the testing situation. Thus, several measures of test-taking behavior have been examined. These measures have the added advantage of providing information on the child's behavior that can be used in interpretation of the test results. The Circus Behavior Inventory, the Stanford-Binet Face Sheet, and the High/Scope Pupil Observation Checklist were evaluated. A comparison of all three (see Table 1) indicated comparable psychometric properties, but the Pupil Observation Checklist had the advantage of having been used previously in a national evaluation and so it was chosen. The POCL is a rating scale of nine 7-point bipolar adjectives developed for the Home Start evaluation from a 25-item scale used by High/Scope in Follow Through. Two factors labeled "Test Orientation" and "Sociability" were found consistently across several time points in the Home Start evaluation. A modification of the instrument incorporating dimensions found in the Stanford-Binet Face Sheet will be developed.

A rating scale to be used by teachers, called the Child Rating Scale, incorporating categories of behavior parallel to the observation instrument, will be developed for use in conjunction with classroom observations and the tester's POCL ratings. Teachers are often considered to be the best judge of their pupils, although they are not always unbiased observers. Thus, disagreement between the Child Rating Scale and the other two instruments would not invalidate the observations or tester ratings, but agreement between the measures would tend to confirm their validity.

Social problem-solving, including self-assertion. This fourth category of social-emotional competence involves the child's desire and attempt to solve social problems, the learning of pro-social as opposed to anti-social roles, and the developing of a range of pro-social (that is, neither violent nor submissive) alternative solutions to real-life interpersonal problems. This includes the child's learning to get his/her way appropriately in a group without being a bully, developing the ability to assert one's right to fair play, and developing the ability to gain access to others, including getting and maintaining the attention of adults in socially acceptable ways.

This category has perhaps the clearest relation to what is commonly thought of as social competence, and is integral to Bronson's definition of social executive skill, which she saw as "the ability to control and direct oneself adequately and constructively in social situations and the ability to influence others effectively in socially approved ways" (1975). It does not imply mere compliance, but rather the effective use of

coping strategies. Charles Billings[1], has stressed the importance of this cateogry, emphasizing that the child's learning to control his or her environment and to influence others was an essential component of social competence and a necessary precursor to the use of social and political power, which is mandatory for an individual's or group's effective functioning in a democracy.

It was decided that the assessment of a child's social problem-solving skills would be done partly by observations, partly by the Child Rating Scale, and partly by the Preschool Interpersonal Problem-Solving Test (PIPS) developed by Spivak and Shure (1974). This test is designed to assess a child's ability to generate alternative solutions to two real-life kinds of problems, the first involving ways for a child to obtain a toy from another child, and the second involving ways to avert the mother's anger which might result from damage to property. Since the two parts of the PIPS test are highly intercorrelated and together takes 20 to 25 minutes to complete, only the first half of the test will be administered. This instrument has high face validity and, according to the Rand report, "is eminently suited for assessing the extent to which Head Start increases prosocial behavioral alternatives" (Raizen and Bobrow, 1974).

Recognition of feelings in self and others, and sensitivity to and understanding of others. The fifth category of social-emotional competence involves the development in the child of awareness of his/her feelings, both positive and negative, and the ability to express them appropriately. It also involves perceiving the emotions of others and responding to them appropriately, perceiving and accepting differences between oneself and others, tolerating others' viewpoints, forming close relationships with peers, and responding positively to handicapped children. (This latter is directly relevant to the PDC goal of mainstreaming the handicapped.)

As with self-concept and learning how to learn, the problems of lack of agreed-upon theory and definition of "recognition of feeling" hindered the process of selecting appropriate instruments for preschool children. "Empathy" may be a related concept, but it was excluded from the Harvard Preschool Project's construct of social competence because it did not differentiate between children rated competent and noncompetent. However, recognizing one's feelings and those of others should be distinguished from empathy, which is usually defined as "taking the point of view of the other," and which may be confounded with egocentrism in preschoolers (Borke, 1972; Chandler and Greenspan, 1972). One measure of "empathy" does exist for preschool children (Feshbach and Feshbach, 1969), but administration procedures calling for the use of slide projectors and the verbatim recording of child's responses make it impractical for a national evaluation.

---

[1]PDC evaluation Advisory Panel meeting, April 1975.

Because of the difficulties in assessing this construct in preschool children, it was decided to assess it by classroom observations and teacher ratings.

Positive school attitude. The final category of social-emotional competence includes, according to Anderson and Messick the child having a positive attitude toward some things in school and not having a generalized negative attitude toward most school experiences. The "socially competent" child feels comfortable in school, not threatened or anxious, and likes to come to school. He/she talks freely to all children and staff in school, not just his/her own teacher and children of his/her own age level.

The discussion of attitudinal constructs in the Rand report clearly sets forth the hazardous nature of attitude assessment because of the tenuous relationship between attitudes and behavior. Furthermore, as Walker warns, there is a "dearth of instruments suitable for young children" (Walker, 1973). Consequently it is expected that archival data, such as number and kind of absences, number of visits to the school nurse, tardinesses, and disciplinary incidents, might form the basis of attitude assessment. Even if the relationship between these behavior indicators and attitudes toward school is not completely straightforward, these behaviors are of more consequence to the school and to the child than are the attitudes.

Because it is important to attempt to assess positive school attitude, another technique is being recommended for pilot testing. A brief self-report instrument, which is a segment of the High/Scope Child Interview used this year in a Follow Through evaluation, is available. In an interview format the child is shown a sheet of paper with five faces on it depicting the range from happiness to sadness. The child is asked to point to the face that shows how he/she will feel about school next year, how his/her teacher feels about the child, and how the child feels about the teacher or teachers. Bogatz and Ball (1971) recommend caution in assuming that disadvantaged preschool children can recognize common emotions, but this measure includes two practice items to ensure that the child understands the meaning of "happy" and "sad".

Alternatives to Testing

The review of instruments available for measuring social and emotional development has been singularly unrewarding but not surprising in view of the lack of social and affective developmental theory. Moreover, as Bronson (1975) has commented, "it can be argued that the current theoretical 'state of the art' in social and emotional development is so primitive that observation may give the best chance to develop a reasonable theoretical approach." Walker (1973) concluded, "The most reliable and valid measures (of socio-emotional variables) available at the present time are

the observational, non-verbal techniques." Therefore, it was decided to place considerable weight on classroom observation for assessing several aspects of social-emotional competence, and to complement the observations with teacher and tester ratings.

Developmental Continuity observation system. Although many researchers have pointed out the drawbacks of naturalistic observations (e.g., Bronson, 1975; Herbert, 1970; Wright, 1960), the advantages to be gained in terms of specificity of naturally occurring behaviors, the lessening of inferences required, and the elimination of artificial and inappropriate demands on the child far outweigh the disadvantages. In this section, the initial developments of an observation system is described. The system will be tried out in Year II (1975-76) and refined for use in 1976-77. At that time a manual detailing the behavior categories and the observation procedures will be published.

The choice of categories to use in the observation system was guided by a number of considerations. The first concern, of course, was that the behaviors observed reflect the social-emotional goals of Project Developmental Continuity. Second, those behaviors which had been found to differentiate "competent" from "incompetent" children in the preschool and early elementary years in the Harvard Preschool Project were examined (White, et al., 1969; White and Watts, 1973; Ogilvie and Shapiro, 1973; Bronson, 1973, 1975; Ross and Zimiles, 1974). Ogilvie and Shapiro's (1973) Social Behavior Checklist and Bronson's (1975) Social and Non-Social Executive Skill Profile contain many child-adult and child-child interaction categories which appear to assess those areas of social-emotional development important to PDC.

Another consideration was the identification of frequently occurring behaviors. High/Scope Foundation staff who have been involved in observing and working with preschool- and early elementary-aged children provided feedback about relevant behaviors which PDC observers could expect to encounter frequently. Because observers will only be in the classroom for a limited amount of time, it was agreed that it would be impractical to train them to reliably observe behaviors which might seldom occur during their observation periods.

Additional considerations were the experience and skills of the individuals being trained to use the observation system. It was expected that the observers would be paraprofessionals accustomed to working with children but unfamiliar with the techniques of systematic observation. Therefore objective, operationally defined categories were chosen rather than categories which depended upon observers' inferences. Further, with only one week available in which to train observers, it was decided that broader, conceptually based categories would be more appropriate and yeild greater interrater reliabilities than narrowly defined categories which required fine discriminations.

Based on the above considerations, behavioral categories were formed by refining or combining categories from existing observational systems, and by adding other behaviors appropriate to PDC goals (the categories are listed in Figure 6). It is expected that the resulting observational system is one which relatively inexperienced observers can be taught to use reliably within a short period of time, and which still addresses the social-emotional goals which the PDC evaluation is interested in assessing.

As mentioned above, teacher ratings (using a set of items designated as the Child Rating Scale) will be used in conjunction with the observation system. Teachers will be asked to rate each child using categories relevant to social competence which are based on Bronson's (1973) Task and Social Competence Rating Scale. Although the problem of teacher bias in the ratings is again unavoidable, to the extent that teacher-observer agreement exists on categories rated by both, some credibility could be attached to those categories rated only by the teachers.

Summary of Recommended Measures for the Basic Battery

The measures recommended for assessing impact in the social-emotional areas are summarized here according to their methodology. The measures are listed with a brief summary statement of the construct being measured.

Child Tests or Interviews

- Child Interview
  Attitude toward school
- Preschool Interpersonal Problem-Solving Test
  Social problem-solving
  Recognition of feelings
- Stephens-Delys Reinforcement Contingency Interview
  Self-directing; locus of control

Teacher or Tester Ratings

- Pupil Observation Checklist
  Sociability and task orientation
- Child Rating Scale
  Categories parallel to the observation system

Direct Observation

- PDC Classroom Observation System
  Feelings of individual worth
  Learning how to learn
  Recognition of feelings in self and others
  (see Figure 6 for detailed list of categories)

Figure 6

Recommended Categories for the PDC Classroom Observation System

1. Non-Involvement

2. Involvement
   a. Social
   b. Non-social
   c. Combined social and non-social

   d. Active
   e. Passive

   f. Self-initiated or structured
   g. Other-initiated or structured
   h. Mutually initiated or structured

   i. Gross motor
   j. Fine motor
   k. Non-motor

   l. Verbal
   m. Non-verbal

3. Interacts with Peers (Active)--Positive or Neutral
   a. Cooperative
   b. Controlling

4. Interacts with Peers (Active)--Negative

5. Imitates Peers (Passive)

6. Uses Peer as Resource

7. Interacts with Adults--Positive or Neutral
   a. Cooperative
   b. Controlling

8. Interacts with Adults--Negative

9. Imitates Adults

10. Uses Adult as Resource

11. Role Play
    a. Solitary
    b. Cooperative

12. Pride in Self
    a. Physical attributes of possessions
    b. Accomplishments or products

Measures of Psychomotor Development, Health and Nutrition

Defining Areas of Psychomotor Competence

There is considerable disagreement concerning the relation of psychomotor skills to social competence. Burton White (1973), for example, has found only "quite modest" differences in motor and sensory capacities between children with high overall competence and children with low overall competence. Since the development of psychomotor skills appears to be more influenced by maturation than are other domains of social competence, it may not be significantly affected by PDC experience, and the likelihood of finding program effects has been questioned by some authorities.

Others believe that the five areas of psychomotor skills outlined below (see also Figure 2) are important components of social competence. One reason for their importance may be that a child lacking in some of these abilities could be differentially perceived by teachers and peers and be at a relative disadvantage in mastering classroom-related skills; this, in turn, may affect later school performance.

The five areas of psychomotor skills identified by Anderson and Messick as part of social competence are fine motor dexterity, gross motor skills, visual and auditory perception skills, and perceptual motor skills. Fine motor dexterity involves the ability to manipulate small objects. Gross motor skills involve large muscle use for activities such as jumping and balancing. Visual perception involves discriminating between similar forms, and auditory perception involves discriminating between similar sounds; both skills are thought to be important in developing competency in language and reading. Perceptual motor skills involve coordinating visual, auditory, and motor behavior in such activities as copying geometric forms and imitating gestures.

Psychomotor instruments were examined within each of the five subcategories and judged for their appropriateness (see Table 2). The most widely used and suitable measures (with selected references) are as follows:

- Purdue Perceptual Motor Survey (Jamison, 1972)
- Denver Developmental Screening Test (DDST)
- Developmental Tasks of Visual Motor Integration (VMI) (Chissom, 1972)
- Gesell gross motor tasks (Werner, 1965)

Table 2

Ratings of Measures of Psychomotor Development on Criteria for Measures Selection

Measure	Practical Considerations					Psychometric Qualities				Relevance to PDC				Past Use
	Available for use by fall 1975	Appropriate for trained paraprofessionals	Test format appropriate for PDC age group	Scoring procedures appropriate for data processing	Reasonable testing time for young children	Adequate construct and/or predictive validity	Adequate test stability and internal consistency	Culture and/or SES fair	Representativeness of standardization sample	Spans age range of PDC population (4-8)	Spanish adaptation available	Relevant to PDC social-emotional goals	Likely to demonstrate program effects	Used in previous national evaluation or large-scale studies
Purdue Perceptual Motor Survey	X	X	X	X	X	X	X	X	X			X		
Denver Developmental Screening Test (Modified)	X	X	X	X	X	X	X	X	X			X		X
Gesell gross motor tasks	X	X	X	X	X	X	X	X				X		
Copy What You See (CIRCUS)	X	X	X	X	X		X					X		
Auditory Discrimination Test (Wepman)	X	X		X	X	X	X					X		
Children's Auditory Discrimination Inventory	X	X		X	X	X	X		X			X		
McCarthy Scales of Children's Abilities	X	X	X	X	X	X	X	X	X	X		X	X	
WPPSI and WISC Block Design Task	X	X	X	X	X	X	X	X	X	X		X	X	

- Copy What You See from CIRCUS
- Auditory Discrimination Test (DiCarlo, 1965)
- Children's Auditory Discrimination Inventory
- McCarthy Scales of Children's Abilities
- WPPSI and WISC lock esign ask

Input was sought from each of the individual PDC sites concerning the importance their program was attaching to psychomotor goals. None of the psychomotor goals was ranked among the most important child goals, yet no program was willing to give them especially low priority.

In view of the possibility that an educational program might have little influence on psychomotor development, the relatively neutral local site evaluation of psychomotor goals, and the narrow scope of each psychomotor instrument, all were rejected in favor of subtests of the McCarthy Scales of Children's Abilities and the Block Design Task of the WPPSI and WISC.

The MSCA is a relatively new test (McCarthy, 1972), with subtests for assessing perceptual-performance and motor development. A number of factors recommended the use of the MSCA: the age range (from 2-1/2 to 8-1/2 years), making it feasible for use in a longitudinal evaluation; its psychometric development, which appeared adequate; its standardization, which was carried out on 1,032 children proportionately representative of the total population in terms of sex, race, geographic location, and SES levels; the apparent ease with which a paraprofessional can administer it; and its integration of psychomotor and cognitive-language tasks. (Discussions of the cognitive and language subtests appear in the next section.)

Perceptual performance. Subtests from the Perceptual Performance Scale and from the Motor Scale were selected on the basis of appropriateness for PDC goals. The Perceptual Performance Scale, "consisting of 'gamelike' tasks which do not require the child to speak, assesses his reasoning ability through the manipulation of materials" (McCarthy, 1972). Both social and non-social competence scores on Bronson's Executive Skill Profile showed significant correlations with this scale of the MSCA. The subtests (numbered in order of presentation in the MSCA) recommended for the basic battery are:

Subtest		Description
1.	Block Building	Child copies block structures built by the examiner.
13.	Draw-A-Child	Child draws a picture of a child of the same sex.
18.	Conceptual Grouping	Child classifies blocks on the basis of size, color, and shape.

Three subtests were eliminated from consideration at the outset: Right-Left Orientation, because it was not recommended for children under five; Puzzle-Solving, because the WPPSI and WISC Block Design Task was thought to be a more powerful measure encompassing the same skill; and the xylophone Tapping Sequence, because it appeared to tap memory more than psychomotor skills. Draw-A-Design (Subtest 12) was deleted after the complete battery was reviewed and it was necessary to reduce testing time; Draw-A-Design was seen as expendable in view of likely redundancy with Draw-A-Child and Block Design subtests.

Block design. The WPPSI and WISC Block Design subtest is recommended to assess problem-solving strategies and conceptual maturity. It is mentioned here since it requires the manipulation of small blocks and may be considered appropriate for the perceptual motor area.

Motor performance. The Motor Scale of the MSCA assesses the child's coordination as he/she performs a variety of gross and fine motor tasks. The recommended subtests are:

Subtest		Description
9.	Leg Coordination	Child performs motor tasks which involve the lower extremities, such as walking backwards or standing on one foot.
10.	Arm Coordination	Child bounces a rubber ball (Part I), catches a beanbag (Part II), and throws a beanbag through a hole in a target (Part III).
13.	Draw-A-Child	Child draws a picture of a child of the same sex.

The first two subtests measure gross motor ability; the last measures fine motor coordination of the hand and fingers. It is recognized that the drawing task involves a strong cognitive component as well, and that it has also been used clinically to assess the child's personality. This is not the purpose of its inclusion here, however, and McCarthy's short, objective, non-inferential scoring system will be used rather than an in-depth interpretative analysis.

Visual and auditory perception. No auditory discrimination or visual perception measure is recommended for the basic battery Since visual perception is involved in the MSCA Perceptual Performance scale and the WPPSI Block Design test, a test specifically for visual perception does not seem warranted, and auditory discrimination does not appear to be of concern to most PDC sites. To accommodate sites that are more concerned with this psychomotor area, the Wepman Auditory Discrimination Task will be made available as a site-specific option.

Defining Areas of Health and Nutrition

The problems of malnutrition and poor health have been a major focus of Head Start from its inception; moreover, recent data have documented the relationship between good health and the child's potential cognitive development (Birch and Gussow, 1970). The PDC Guidelines explicitly require that each site assess the nutritional, medical, dental, and social service needs of PDC children, locate community resources, and develop a system to meet these needs. Head Start requires medical, dental, and speech diagnostic screenings of a prescribed nature as well as the completion of all recommended immunizations and dental treatment.

Recommended procedures. As a consequence of the specificity and concreteness of the performance standards, it is recommended that the Office of Child Development provide guidelines for standardized record-keeping so that child health information can be gathered uniformly from all sites. Responses to the goals questionnaire reflected a strong emphasis on health and nutrition across all sites and an endorsement of the goal of the child's attaining not only a minimally acceptable criterion of health, but reaching a level of "positive" health. In order to assess these goals, the following measures are recommended, grouped according to goal and level of attainment:

- Absence of disease, freedom from correctable handicap

  Records of immunizations (rubella, rubeola, mumps, DPT, polio)
  TB exam
  Dental exam for caries

Visual exam: Snellen and/or directional letters test
Audiometric screening test for hearing loss

- Level of positive health

  Height, weight
  Observation of amount of energy, vigor
  Frequency of absences for health reasons
  Sicknesses at school, visits to nurse
  Iron deficiency tests (hematocrit, serum ferritin)
  Protein deficiency test (serum albumin)

- Better eating habits

  Sample school menus

Summary of Recommended Measures for the Basic Battery

- Psychomotor Development

  Perceptual Performance

  Block Building (MSCA)
  Draw-A-Child (MSCA)
  Conceptual Grouping (MSCA)

  Motor Performance

  Leg Coordination (MSCA)
  Arm Coordination (MSCA)
  Block Design (WISC/WPPSI)

- Health and Nutrition

  Head Start, Program and School Records

  Records of immunizations
  Incidence of TB
  Iron deficiency tests
  Protein deficiency test
  Dental exam
  Snellen test and/or directional letters measure
  Audiometric screening
  Height
  Weight
  Vigor measure
  Frequency of absences
  Sicknesses at school
  Sample school menus

Measures of Cognitive and Language Development

Traditionally, national educational evaluations have emphasized cognitive outcomes as the major criteria for program success. In contrast, the PDC evaluation views cognitive and language abilities as only part of many factors which comprise the child's "everyday effectiveness in dealing with his/her environment." Such a perspective acknowledges the acquisition of cognitive abilities as being interrelated with and necessary for the child's psychomotor, social-emotional and language development, but de-emphasizes cognitive growth as the major goal independent of other goals.

Defining Areas of Cognitive and Language Competence

The precise nature of the cognitive and language abilities which a child must possess in order to be "socially competent" continues to be debated. In part, controversy exists because the essence of these competencies varies with chronological age and situational contexts. It is clear, however, that across all age levels, to be socially competent within the school environment the child must acquire a set of skills which facilitate academic learning appropriate for his/her developmental level. These competencies are known to most educators, and are represented within Anderson and Messick's (1974) list of social competencies. They include:

- Attentional skills
- Classification/categorization skills
- Perceptual skills
- Information-processing skills
- Evaluative skills
- Memory skills
- Basic language skills

Although this list is clearly not comprehensive, the acquisition of these skills may potentially predict academic performance or the ability to learn concepts and, in turn, social competency in the school environment. Moreover, it may be argued that such skills are inherent in social competency in the home or community environments as well. The child's ability to solve practical problems (e.g., fixing a broken toy) may in part be dependent on

his/her ability to use skills deemed necessary for academic performance. For example, the child must first employ attentional and perceptual skills in order to perceive the problem, classification or categorization skills in order to define and communicate the problem, or evaluative skills for determining the most appropriate solution. This suggests that an overlap exists among skills essential for academic performance on the one hand and practical problem-solving on the other.

While there appears to be some relationship between proficiencies which predict ability to learn concepts and ability to solve practical problems, a stronger relationship emerges between cognitive abilities and communication and/or language competency. In order to effectively communicate relevant observations about his/her environment, the child must direct his/her attention to salient cues, detect similarities or discrepancies with existing knowledge, organize them logically and possess an adequate vocabulary and syntax to express the observation verbally. In addition, there is a whole array of communication competencies (e.g., lack of egocentricism, taking the listener's point of view) which strengthen the existing interplay among language and cognitive abilities.

The complex relationship between cognition and communication suggests that an appropriate approach for measurement would be the assessment of cognitive and language abilities that serve as the basis for academic achievement, practical problem-solving and verbal competency. Such an approach receives further support from PDC staff and teachers who designated the goals of "problem-solving" and "general language use" as important and indicated they were emphasized by their programs. This suggests that at least some PDC classrooms attempt to provide an environment where problem-solving and general language competencies are nurtured, and, consequently, program effects may be observed when PDC and non-PDC children are compared. In light of these theoretical and program issues, a brief description of the three major cognitive and language areas is offered:

General cognitive competencies. This area encompasses the child's age-related ability to attend to salient cues of stimuli; perceive units or forms as separate from their context; synthesize units into an organized form; apprehend the nature of similarities and differences; categorize objects or events according to their attributes; and use strategies of information processing for appropriate retrieval.

Problem-solving competencies. This refers to the child's age-related ability to apply general cognitive competencies in order to identify the proper solution for a problem and to evaluate his/her own responses and products in the process.

Verbal competency. This area relates to the child's age-related abilities to verbally label objects and events appropriately; describe objects and events speaking audibly and comprehensibly; repeat information told to him/her; and make requests or give instructions.

Measurement Selection Process

The selection process began with the review of cognitive and language measures utilized in other national evaluations of early childhood programs, those recommended by the Rand report, and those suggested by the panel of High/Scope consultants. This review process yielded a list of possible measures to undergo further examination for PDC relevancy. Table 3 presents the list of measures rated according to the 15 criteria employed for determining their suitability for use in the PDC evaluation.

The ratings in Table 3 make it evident that tests employed in previous evaluations fail to meet several of the criteria. The primary shortcomings are that many of the tests do not span the total age range (4 to 8 years) of the PDC population, are irrelevant to PDC cognitive and language goals, do not fully represent geographic or SES aspects of the PDC population in their standardization procedures, or have been found to be highly correlated with general ability measures. Thus, the criteria employed to select tests with the most desirable characteristics considerably reduced the pool of potential measures for the PDC battery.

Given the inadequacies of many developed measures, and considering the constraints on developing new measures, a reasonable course of action is to propose the use of carefully selected subtests extracted from the available cognitive and language measures. By examining a variety of subtests it is possible to select ones that more fully reflect the desired characteristics of the PDC measurement battery.

Because of the dangers inherent in separating subtests from the context in which they were developed, several precautions have been taken. First, the tests were examined for desirable psychometric characteristic directing particular attention to elements of the individual subtests. The subtests under consideration from the WPPSI and WISC, all displayed acceptable internal consistency or stability coefficients and concurrent or predictive validity independent of the total measures (see appendix). However, because the validity and reliability coefficients of the MSCA subtests are neither published nor available at this

Table 3

Ratings of Measures of Cognitive and Language Development on Criteria for Measures Selection

Measure	Practical Considerations					Psychometric Qualities					Relevance to PDC				Past Use
	Available for use by fall 1975	Appropriate for trained paraprofessionals	Test format appropriate for PDC age group	Scoring procedures appropriate for data processing	Reasonable testing time for young children	Adequate construct and/or predictive validity	Adequate test stability and internal consistency	Culture and/or SES fair	Representativeness of standardization sample	Low correlation with index of general information	Spans age range of PDC population (4-8)	Spanish adaptation available	Relevant to PDC cognitive language goals	Likely to demonstrate program effects	Used in previous national evaluation or large-scale studies
CIRCUS	X	X	X	X	X		X				X		X	X	
Cognitive Abilities Test	X	X		X		X	X		X						
Cooperative Primary Test	X	X		X			X								
Coloured Progressive Matrices	X	X		X	X	X	X	X	X	X		X	X		X
ITPA	X	X	X	X			X				X		X	X	X
McCarthy Scales of Children's Abilities	X	X	X	X	X	X	X	?	X		X	X	X	X	
Peabody Picture Vocabulary Test	X	X	X	X	X	X	X							X	X
Preschool Inventory	X	X	X	X	X	X	X							X	X
Stanford-Binet	X	X	X	X	X	X	X							?	X

Continued:

Table 3 (continued)

Ratings of Measures of Cognitive and Language Development on Criteria for Measures Selection

Measure	Practical Considerations					Psychometric Qualities					Relevance to PDC				Past Use
	Available for use by fall 1975	Appropriate for trained paraprofessionals	Test format appropriate for PDC age group	Scoring procedures appropriate for data processing	Reasonable testing time for young children	Adequate construct and/or predictive validity	Adequate test stability and internal consistency	Culture and/or SES fair	Representativeness of standardization sample	Low correlation with index of general information	Spans age range of PDC population (4-8)	Spanish adaptation available	Relevant to PDC cognitive language goals	Likely to demonstrate [illegible]	Used in previous national evaluation or large-scale studies
WISC	X	X	X	X	X	X	X	?	X	X		X	X		
WPPSI	X	X	X	X	X	X	X	?	X	X		X	X		

Table 4

Summary of Measures for Cognitive and Language Basic Battery

Subtest	Test	Stimulus Materials	Child's Task	Skills Assessed	Administration Time	Administered to Bilingual Children
Conceptual Grouping	MSCA	Set of 12 blocks: 6 squares, 6 circles, each shape in three colors and two sizes	To perform various manipulations along the 3 dimensions of color, size and shape according to verbally presented instructions	Classification Categorization Perception Attention Understanding, size, color, shape	10 minutes	Yes, with modification
Verbal Memory	MSCA	Verbal presentation of sequence of 3 and 4 words, 14-word sentences and a one paragraph story	To repeat as much of the stimulus presented as possible	Memory Information-processing Attention	20 minutes	No
Block Design	WPPSI	14 flat blocks, solid red on one side, 6 blocks solid white on the other side, 8 blocks half red, half white (diagonally) on the other side	To construct the appropriate design either from examiner's demonstration model or from model design pictured on cards within fixed time limits for each design	Problem solving Attention Perception Evaluation Classification	10 minutes	Yes
Block Design	WISC	9 cubes, 4 solid color faces (blue, yellow, red, white), two diagonally split faces (red/white, yellow/blue)	same as above	same as above	10 minutes	Yes
Verbal Fluency	MSCA	Verbal request for members of a specific category	To name as many subordinate members as possible within 20-second interval	Classification Categorization Memory Evaluation Verbal expression	10 minutes	Yes
Opposite Analogies	MSCA	Verbal presentation of simple analogies	To provide the missing word in an analogy	Memory Attention Classification Categorization Evaluation	7 minutes	No

time, some uncertainty surrounds the characteristics of subtests apart from the total MSCA subscales. Therefore, unlike the WPPSI and WISC, selection of the MSCA subtests could not be based on psychometric indicators, but occurred within the framework of social competence developed for this study. Further guidance was provided by a Head Start evaluation study (Rentfrow, et al., 1972) which also extracted subtests from the MSCA to comprise a measurement battery. The results of this evaluation (N=64) indicated that the subtests preferred for PDC correlated moderately (r=.57 to .70) with the composite score. Evidence on the MSCA factor structure is available from Kaufman (1973) and Kaufman and Hollenbeck (1973).

Since so many of the tests initially considered were found to favor a specific culture, racial or socioeconomic group, a second precautionary measure was an extensive review of the literature and discussions with bilingual/bicultural consultants. These procedures were specifically enacted to identify which of the preferred subtests exemplify cultural and socioeconomic "fairness". Because of the importance and relevancy of this factor for the PDC evaluation battery, this "fairness" element was weighted heavily in making the final recommendations. Information regarding the influence of culture and socioeconomic status on WISC/WPPSI and MSCA test performance were obtained from various independent small scale studies.

Kaufman and Kaufman (1973), for example, designed and directed a study in order to determine differences in Black and White performances on the MSCA, a critical study since only limited data were provided by the standardization study. This study indicated performance differences, favoring the White child on the Word Knowledge and Number Questions subtests with older children (6- to 8-1/2 years). Since these subtests appeared to be highly dependent on learned concepts or academic attainment for the older levels, it was decided to exclude any subtest that appeared to tap these elements.

MSCA subtests. The McCarthy scales (McCarthy, 1972) were developed to assess a variety of cognitive and motor behaviors of children from 2-1/2 through 8-1/2 years of age. The content of the tasks was designed to be suitable for both sexes as well as for children from a variety of ethnic, regional, and socio-economic groups. The materials and questions are administered individually in a gamelike manner. All subtests are easily and quickly administered to avoid taxing the young child's attention span. Taking account of these considerations and those discussed

earlier, the following MSCA subtests are recommended for the basic battery of cognitive and language measures:

- Verbal Memory
- Conceptual Grouping
- Opposite Analogies
- Verbal Fluency

Each of these subtests is described in Table 4, with information on the stimulus materials used, a brief description of the child's task, the skills assessed, administration time and appropriateness for bilingual children.

Block Design. The WPPSI and WISC scales (Wechsler, 1949 and 1967) represent a downward extension of the Wechsler Adult Intelligence Scales. Since the WPPSI is appropriate for ages 4 to 6 and the WISC is appropriate from age 5 through adolescence, subtests from both can be used to accommodate the age range of PDC. The WPPSI and WISC, typically used in clinical evaluations to calculate IQ, consist of 11 individually administered subtests to assess verbal and cognitive performance. The materials were designed to appear intrinsically interesting to the child and have been used widely across socioeconomic, ethnic and regional groups.

Considerable research has been conducted on the WISC and WPPSI (e.g., Littell, 1960; Woo-Sam and Zimmerman, 1973; Zimmerman and Woo-Sam, 1972), but there is little information on the Block Design subtest as apart from the scale score. The Block Design subtest does appear to be highly correlated with full-scale IQ, but somewhat more highly related to performance IQ than to verbal IQ. The fact that it is easily administered, brief, nonverbal and highly reliable suggests its suitability for inclusion in a battery that is attempting to assess a wide range of competencies.

As a stable measure of cognitive problem-solving at all age levels, the Block Design subtest can serve as a useful covariate and as an entry-level measure of PDC-comparison group differences.

Summary of Recommended Measures for the Basic Battery

The subtests recommended are grouped according to three general dimensions of competence:

General Cognitive Competence

Conceptual Grouping (MSCA)
Verbal Memory (MSCA)

Problem-Solving Competence

Block Design (WPPSI, WISC)

Language Competence

Verbal Fluency (MSCA)
Opposite Analogies (MSCA)

Measures of Impact on Parents and Teachers

Head Start philosophy has always recognized the significant roles parents and teachers play in affecting children's social competence. Various Head Start and Follow Through models have actively encouraged parents' participation in school decision-making. Numerous efforts by Head Start Planned Variation and Follow Through sponsors have been undertaken to help teachers increase their awareness of the particular needs of the lower socioeconomic, bilingual, and handicapped child and family. OCD has designed goals for PDC which are consistent with those of previous Head Start and Follow Through projects, emphasizing joint parent and teacher participation in defining educational goals and making educational decisions. These decisions will be primarily concerned with the key PDC requirement that the Head Start and elementary school teachers develop a coherent and continuous curriculum based on the developmental needs of children from Head Start through the third grade. The purpose of the evaluation will be to assess the impact of PDC on parents, staff and teachers in terms of attitude and behavior change, as well as subsequent changes in the school environment.

Defining Program Goals for Parents and Teachers

The PDC Guidelines deal with parent involvement as a major component of the program and also specify some issues in the area of staff development. An examination of these Guidelines and discussions with local and national PDC staff led to a definition of outcomes in the following areas:

Parent outcomes:

- Participates in school and classroom activities. The parent perceives his/her role in school activities and processes as providing the necessary bridge between home and school. This perception is translated into an active involvement as a classroom volunteer, room helper, lunch room supervisor, volunteer tutor or other voluntary school roles.

- Demonstrates understanding of a child's developmental educational process. The parent recognizes the acquisition of knowledge regarding his/her child's socioemo[illegible], cognitive, psychomotor and academic growth as enhancing his/her role as "facilitator of development." This recognition may be evident in the types of experiences and materials available in the home or a possession of realistic expectations and aspirations for their child's academic achievement.

- Provides input into school decision-making and problem-solving activities. The parent perceives his/her role in school decision-making and problem-solving activities as essential for reflecting community and family needs. This perception translates into observable input into group decision-making or problem-solving processes during school meetings.

- Increases personal development through participation as a decision-maker and problem-solver. The parent is aware that effectiveness in group processes is based on his/her acquired abilities as a competent decision-maker or problem-solver. In recognition of this contingency, the parent takes full advantage of training sessions or other group meetings for the enhancement of these qualities.

Teacher outcomes:

- Provides instruction matching the child's developmental learning level. The teacher perceives the adoption of a developmental learning perspective as essential for promoting socioemotional, cognitive, psychomotor and academic growth. This awareness is operationalized through individualization of instruction to match the child's learning level. The teacher demonstrates competencies in assessing this learning level and maintaining current profiles of each child's learning abilities.

- Creates a classroom environment conducive to the development of social competency. The teacher perceives attendance at child development inservice training sessions or discussions as essential to learning how to create an appropriate climate for nurturing social competency. The teacher then takes tangible steps to create such a climate by adopting a warm, child-directed interaction style, providing a balance between child-initiated and teacher-initiated activities, utilizing materials appropriate for given developmental levels, encouraging curiosity and discovery behaviors, and helping the child to obtain skills necessary for social problem-solving competencies.

- Acquires knowledge of resources available to meet needs of PDC children and families. The teacher is aware that to provide the best support services for children and families, there must be a knowledge of available community and school district services as well as an awareness of possible roles parents might play as assisting agents. This awareness is reflected through active participation in support services training sessions where such information may be obtained and discussed. In addition, the teacher acquires skills for interacting with and instructing exceptional children in the classroom.

- Provides activities for bilingual/bicultural children. The teacher perceives an active participatory role in bilingual/bicultural inservice training sessions as increasing his/her awareness of the needs of bilingual/bicultural children and families. This perception is manifested in utilization of multi-cultural materials, relevant learning experiences and assisting the bilingual/bicultural child's adjustment into the mainstream of the classroom and school.

Staff-as-a-group outcomes:

- Staff members interact with members of differing grade levels. The teachers as a group share a common recognition that adequate communication, knowledge concerning the developmental nature of a child's learning experience, and a shared philosophy can only be obtained through interaction with each other. Teachers interact across grade levels through discussions, observations in various classrooms, or trading teaching assignments.

- Staff members jointly plan and develop educational and program goals for children, parents and themselves. The teachers as a group perceive cooperative planning efforts as necessary for achieving continuity in an educational and support services-oriented program. This perception results in attendance and participation in staff meetings designated for curriculum materials selection or articulation of program goals.

- Staff members commonly gain competencies in group problem-solving and decision-making techniques. The teachers as a group possess a shared awareness regarding the necessity for a productive group process of acquiring effective group problem-solving and decision-making skills. This awareness is demonstrated by attendance at inservice training sessions or small problem-solving discussions where such skills may be fostered and matured.

Institutional change outcomes:

- Consistent staff, school and parent interaction. Teachers and parents commonly perceive that a shared PDC philosophy can only be implemented through consistent interaction with one another. This perception becomes manifest during parent- or teacher-initiated conferences for cooperatively discussing a given child's status in the classroom and home, joint attendance at relevant training sessions, and joint participation in school or PDC meetings and activities.

- United effort by teachers and parents in designing and developing educational goals. Teachers and parents alike recognize that a PDC program which reflects the needs of children and the community, the child's developmental level, continuous educational experiences, and enhancement of social competency can only be achieved through united planning attempts. This recognition is operationalized with teachers and parents attending and actively participating in planning-oriented committee meetings or goal-setting sessions.

- Integration of Head Start and elementary school philosophies and services. Staff and parents perceive the integration of Head Start and elementary school philosophies and services as necessary for promoting developmentally continuous educational and support services programs. This perception results in a shared PDC philosophy, the adoption of a continuous educational approach, continuous provision of services for children and families, as well as consistent parent involvement, continuing group planning efforts and adequate communication across grade levels and with parents. Both staff and parents possess clearer goals for themselves as individuals, and as a group working cooperatively to achieve a developmentally continuous program. Achieving these goals is equally viewed as rewarding and satisfying by parents and staff.

It should be clearly recognized that many of these goals represent a radical departure from long-established school practice. To include parents in the process of educational decision-making is an innovative procedure. PDC is emphasizing the sensitization of teachers to the special needs of the lower socioeconomic, bilingual, and handicapped child, a process which requires in many cases the development of innovative curricula and new classroom procedures. Another break with established tradition is the requirement that Head Start and elementary school teachers jointly plan a coherent curriculum. To document the progress which the schools make in implementing these changes is an important goal of the evaluation.

In the past teacher and staff support, or the lack thereof, for innovative educational programs originating from the government has received insufficient attention from evaluators, despite its critical importance to the success of these programs. This evaluation is attempting to remedy this defect by the inclusion of measures which will address institutional change in terms of adjustments in parent, teacher, and staff attitudes and interactions.

Selecting Measures for Assessing Impact

A few studies have demonstrated the importance of assessing both attitudes and behavior, although the relationships among them are far from straightforward. Stanford Research Institute's (1971) Head Start study found that a teacher's perception or attitude toward his/her program or sponsor was highly related to child gains on the Wide Range Achievement Test. Coughlan and Cooke (1974) found that in schools where sixth grade children obtained higher reading scores, teachers had more favorable attitudes toward relationships among teachers and toward [illegible] relations, and had higher levels of job satisfaction. Other studies (Klein, 1973; Prescott and Jones, 1972) have found similar attitude-outcome relationships. Findings such as these suggest that an evaluation of teacher outcomes should also include measures of attitudes which may accompany specific behavior change.

Parent attitudes toward school and teachers may also be important indicators of program functioning, but there has been less research on these relationships. In their analysis of Follow Through data, Abt Associates (Cline, 1974) investigated the relationship between parents' views of school "receptivity" and their degree of interaction with teachers or participation in school meetings. Their analysis found no relationship among these variables.

Social psychologists have been studying organizational development for a number of years (Fountain, 1975; Schmuck and Miler, 1971; Schmuck, Runkel, Saturen, Martell, and Derr, 1972). Since many of the early findings were disappointing (Nadler, et al., 1974), recent efforts have been directed toward improving approaches for assessing organizational change. For example, staff of the Survey Research Center of the University of Michigan have been developing models for measuring change in organizations (Robert Cooke and Philip Mirvis, personal communication, 1975). Their work provides a foundation for developing a comprehensive assessment procedure for PDC.

In developing these procedures, several principles have been suggested to guide the measurement and analysis process. First, the measurement process must be multi-level and multivariate; a wide range of variables dealing with individuals, groups and the new entity linking Head Start and elementary school need to be assessed. Second, multiple methods should be employed; the evaluation should include techniques for gathering data which use a variety of methods (observations, archival records, and unobtrusive measures) in addition to self-report instruments (questionnaires, inventories, and structured interviews). Third, it was decided to use a standard set of measures across all sites.

The multi-method approach adopted in this evaluation is expected to generate data allowing statements which will go beyond simple descriptions of outcomes and include reasons for the outcomes. This requires an analysis of the relationships among a variety of variables: program inputs, perceptions, attitudes, and behavior. Figure 7 presents one "model" for conceptualizing the interplay of factors important in the change process. As the reasons for program effects are understood more fully, the evaluation will be able to provide more useful information for program planners and decision-makers to guide the development of future programs.

In addition to facilitating the understanding of these relationships, the assessment of a variety of outcomes provides stronger evidence of program effectiveness. According to some of the research in organizational change, changes in attitudes and feelings of individuals are indicative of a more durable program effect. One might argue that behavioral changes alone (e.g., parent involvement in the classroom) may be shortlived if the parent's feelings about the program are not affected. On the other hand, a change in attitude (e.g., parents say they like the program) might be less valued if it is not accompanied by outward manifestations of the improved attitude.

Summary of Measurement Procedures

The expected outcomes in terms of the behavior of parents, teachers, and staff are listed in Figure 8; Figure 9 lists outcomes in terms of attitudes and perceptions. Beside each outcome is a list of data collection methods that will be employed. The case studies are already dealing with some of these outcomes through the use of structured interviews (one of the self-report methods) and the examination of archival records such as minutes of committee meetings. The primary purpose in this section of the report is to recommend additional self-report measures that will supplement the other procedures and which focus on the attitudes and perceptions of program participants. Among those instruments that have been examined, three have been found which contain items or scales appropriate for assessing the impact of PDC:

- Michigan Assessment of Organizations (Nadler, et al., 1974). During the past two years this instrument has been developed for third party assessment of change in organizations. It has been used in a variety of industrial, service and educational settings. The MAO offers a pool of questionnaire scales and items for measuring individual role perceptions, goal clarity and feedback, climate, group processes and effectiveness, and individual differences.

Figure 7

Model for Viewing PDC Impact

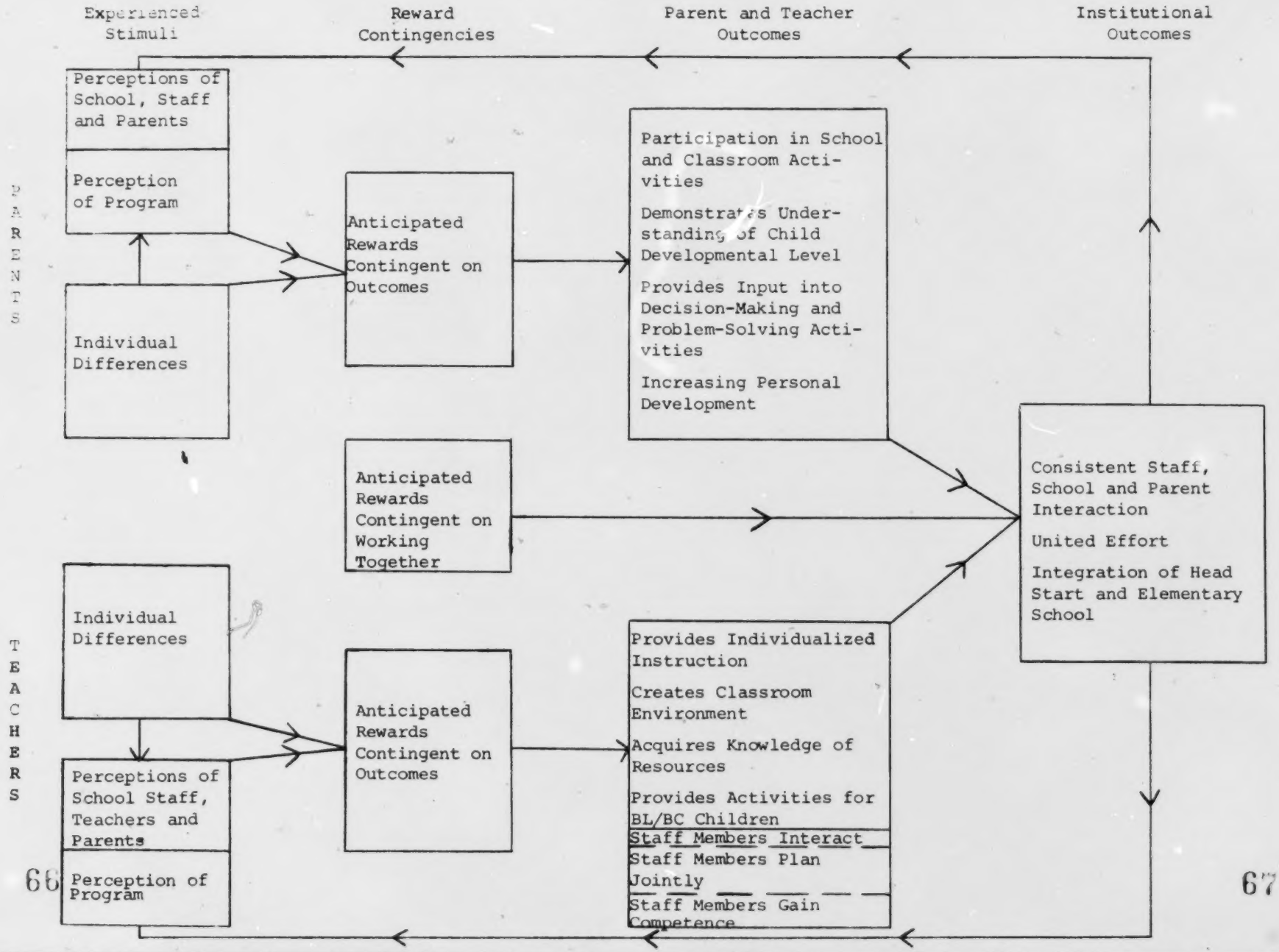

Figure 8

Types of Assessment Procedures for Measuring Impact on Parents, Teachers, Staff, and Institutional Change

PDC PROGRAM OUTCOMES	ASSESSMENT PROCEDURES
Parent Outcomes	
Participation in school and classroom activities	Archival sources Self report Observation
Demonstrates understanding of a child's developmental - educational process	Self report
Provides input into school decision-making and problem-solving activities	Observation Self report Archival sources
Increases personal development through participation as a decision-maker and problem-solver	Self report Archival sources
Teacher Outcomes	
Provides instruction matching child's developmental learning level	Observation Archival sources Unobtrusive measures
Creates classroom environment conducive to development of social competency	Observation
Increases personal development through participation as a decision-maker and problem-solver	Self report Archival sources
Provides activities for bilingual/bicultural children	Observation Unobtrusive measures
Staff-as-a-Group Outcomes	
Staff members interact with members of differing grade levels	Self report Unobtrusive measures
Staff members plan and develop jointly educational and program goals for children, parents and themselves	Archival sources Self report
Staff members commonly gain competencies in group problem-solving and decision-making techniques	Archival sources Self report
Institutional Change Outcomes	
Consistent staff and parent interaction	Archival sources Self report Observation
United effort by teachers and parents in designing and developing educational goals	Archival sources Self report Observation
Integration of Head Start and Elementary school philosophies and services	Self report Archival sources Observation

Figure 9

Instruments for Assessing Perceptions and Attitudes

PERCEPTIONS AND ATTITUDES	INSTRUMENT
Parents	
Perception of program	Interviews Attitude Questionnaire
Perception of school	Attitude Questionnaire
Perception of reward contingencies related to behavior outcomes	Attitude Questionnaire
Staff and Teachers	
Perceptions of school, staff, teacher and parents	Attitude Questionnaire
Perception program	Attitude Questionnaire Interview
Perception of reward contingencies related to behavioral outcomes	Attitude Questionnaire

- The School Survey (Coughlan, 1966). The Survey, used in more than 100 elementary and secondary schools, consists of 120 statements (to which the respondent agrees or disagrees) representing 14 factors related to morale, motivation, and educational effectiveness. The scales were found to discriminate between high and low performance schools and to have good reliability (Coughlan and Cooke, 1974). Scales and items most relevant to PDC include administrative practices, professional and non-professional work load, educational effectiveness, school-community relations, colleague relations, voice in educational program, and supervisory relations.

- The Educational Forces Inventory (Rayder and Body, 1975a, b). This inventory allows the evaluator to "chart" the various forces in the teacher's social-psychological field. The teacher or other staff person is asked to rate the degree of influence 13 forces exert in the classroom process and the degree to which this influence is negative or positive. It has been used in a Follow Through evaluation and possesses good validity (correlated with a measure of teacher morale) and internal consistency. If expanded for coordinators' and administrators' use, this inventory could assess the perceptions of influences from several perspectives.

Sets of items or factors will be selected from these instruments to develop a PDC Teacher Survey. A Parent Survey will also be constructed to assess outcomes for parents.

Bilingual/Bicultural Measures

One of the cornerstones of Project Developmental Continuity is the recognition of individual differences, not only in relation to children, but also, on a broader level, in relation to sites. All of the sites have unique characteristics, but four sites are distinguished by the designation, "Bilingual/Bicultural Demonstration Site." This means that these sites are expected to provide a bilingual/bicultural educational approach, which is described in the PDC Guidelines as using the child's native language for instructional purposes to "reinforce the child's positive self-concept by validating his culture."

The four bilingual/bicultural demonstration sites have a substantial number of bilingual children, those whose "use and comprehension of the English language are limited, and who speak a language other than English at home" (OCD, Guidelines, 1974). The four sites vary in degree of bilingualism from 9% Spanish-speaking children in one Head Start center in Colorado to 100% Navajo-speaking children in Arizona. Several other sites, not designated as demonstration programs, also have a sizeable percentage of bilingual children.

This emphasis on the recognition of the uniqueness of bilingual children is important because of the growing awareness of the essential link between a child's language and his/her identity, and because of the negative outcomes associated with denigrating an integral part of a child's self-concept (United States Commission on Civil Rights, 1975). As Arturo Gutierrez states, "If a child is dominant in Spanish, the use (or lack of use) of that language for instruction, the teacher's attitude toward the use of the language, the teacher's acceptance or rejection of the language, values, total cultural make-up of the child, could have serious negative effects on the child's personality--his self-concept, confidence, his perceptions, etc." (Raizen and Bobrow, 1974b). The PDC Guidelines reflect this concern and insist that "the program must enrich and expand upon the strengths that the children bring to a new learning situation and not follow a compensatory-deficit model... The native language and culture of the children will be regarded as an asset...."

This recognition of the value of individual differences affirms a relatively recent general interest in preserving cultural uniqueness, and reflects social scientists', historians',

and educators' turning away from the traditional esteem in which most had held the "melting pot" theory in favor of prizing cultural heterogeneity (Higham, 1975; Wacker, 1975). It is not enough, however, to affirm cultural uniqueness if the variety that exists within the cultures is not recognized. A major theme running throughout the reaction papers of Rand's Spanish-surnamed panelists is the danger of assuming that there is such a thing as a monolithic Hispanic culture (Raizen and Bobrow, 1974b). In fact there exist significant variations--regional, rural/urban, SES, historical, linguistic, and country-of-origin variations--all but the latter of which are relevant to the three Spanish bilingual/bicultural sites. The California PDC site is an agriculturally based town with most Mexican-American children being second and third generation; the Texas site is a small border town with many first-generation, monolingual Spanish children; and the Colorado site is an industrial city with predominantly monolingual (English-speaking) Mexican-American families who have resided there for many generations. The fourth bilingual/bicultural site, in Arizona, is a low population-density, rural area where most Head Start children are monolingual in Navajo. Obviously these individual BL/BC demonstration sites have unique characteristics which place special constraints on test selection and interpretation.

The other fundamental tenet of this project is the reduction of discontinuity in the child's experience not only between Head Start and the elementary schools, but also between home and school, and, in the case of bilingual children, this requires bilingual education. "Bilingual/bicultural education is designed to help the child make the transition from home to school more easily by reducing the differences between the language and culture of the home and that of the school" (United States Commission on Civil Rights, 1975).

Defining Bilingual/Bicultural Measurement Areas

All of the problems encountered in selecting appropriate measures for English-speaking preschool children were exacerbated in the attempt to find measures appropriate for Spanish-speaking, mixed-language or "Spanglish"-speaking children, and Navajo-speaking children. As part of the PDC measurement design, the BL/BC sites were to receive bilingual adaptations of the basic battery for all sites as well as site-specific bilingual measures.

In appreciation of the fact that testing bilingual/bicultural children presented many complex issues, several sources were consulted. First, very serious consideration was given to the reaction papers of the Spanish-surnamed panelists of the Rand

Corporation to its report on evaluating social competence (1974b). Second, the United States Civil Rights Commission report, A Better Chance to Learn: Bilingual-Bicultural Education (1975) was examined, and finally, recognized experts in the field of bilingual research were consulted. A summary of the main issues which these sources delineated and which were relevant to PDC is presented because they form the basis for selecting, modifying, administering, and interpreting any instruments to be used with bilingual and/or Hispanic children.

One of the most important considerations which stands out when an evaluation of bilingual/bicultural children is conducted is the necessity for specifying the social domains for patterns of language use. This point was emphasized by Luis Laosa in a paper describing some varying patterns of language use in different social domains among different Hispanic groups. "Social domain" is defined as "major spheres of activity in a culture, such as family, education, recreation" (Laosa, 1975). Again the inaccuracy of the assumption that a monolithic culture exists was made clear.

The point was also made by Slobin, who stressed the need for linguists to ascertain the relation between social milieu and the development of language competence, because the effect of the family on language socialization practices was clearly significant but had hitherto been neglected (1967). He admitted that linguists knew very little about children's competence in the use of language, how they learned when to speak and when to remain silent, with whom to use slang and with whom not to, and when to shift languages in a multilingual community.

Laosa's research addressed this issue of the social context of mother-tongue maintenance, in which a stable bilingualism persists, versus the context of language shift, in which the separate domains of languages vanish, and the speaker uses both languages within one sentence. He found that there were significant subcultural variations in language patterns in different social domains between Puerto Ricans in the Bronx, Cubans in Miami, and Mexican-Americans in Austin, Texas. For example, Mexican-American children in Austin were exposed to significantly more English and "Spanglish" (language mixing within the same sentence) in the home and spoke far more English in spontaneous peer interactions in school than either the Cuban American or Puerto Rican children. In a border town like the Texas site, however, the pattern of language use of Mexican-Americans might be quite different because of contiguity with Mexico. It is concluded, therefore, that data on the family and social ecology must be collected and utilized in any evaluation of bilingual/bicultural children.

The specification of language usage in the social domain is important also because there is evidence that a well-developed language system appropriate for each domain may exist (Cruz, in Raizen and Bobrow, 1974b), and therefore many tests in Spanish are inappropriate for the Hispanic child if the child's formal education has taken place in English (Carrow, in Raizen and Bobrow, 1974b), or if it has been normed on Cuban or Puerto Rican children (Cruz, in Raizen and Bobrow, 1974b). Laosa also made the distinction between language dominance and language use. A child may actually use Spanish more, especially in the non-school domain, but be "dominant" (i.e., more literate) in English.

To address this issue of specifying language dominance and social domain Gutierrez recommended the following (Raizen and Bobrow, 1974b):

- Determine the language and cultural dominance initially of all children in the child's classroom--Bernal recommended using language dominance as a covariate (Raizen and Bobrow, 1974b);
- Determine the length of time that one or both languages are used during the day (and, presumably, in which contexts);
- Determine the teacher's emphasis, either on Spanish or English language development, social-emotional development, or other areas;
- Determine what kind of structured language program, if any, has been implemented.

Another major concern of some of Rand's Spanish-surnamed panelists related to the specific language program implemented in the classroom. Gonzalez took the position that the maintenance of the child's first language, assuming that it was Spanish, was of primary importance for the social competence of Spanish-speaking children in the barrio. A maintenance approach is also consistent with the requirements that bilingual/bicultural programs reduce the discontinuity between home and school and encourage the child's positive self-concept. It might also be expected to maintain the child's respect for his parents.

The problem of determining just what constitutes a positive self-concept in each cultural community was raised by Cruz. This is a concern of deep consequence to PDC and is expected to be addressed through the use of indigenous examiners who will also be asked to serve as consultants in many aspects of the testing. This broader use of the indigenous examiners was suggested by Luis Laosa who recommended the following:

- Indigenous examiners should examine the instructions and items on each of the instruments in order to substitute phrases reflecting the regional dialect where "standard" Spanish might be inappropriate.

- Regional modifications should be standardized within each area in advance for use on both pre- and post-tests.

- Objects or toys familiar to children in that locale should be substituted for those that are unfamiliar.

- Indigenous examiners should be used to interpret the scores, to place them in the specific cultural context.

- Indigenous examiners should provide information as to the local language patterns according to social domain.

The issues raised by the Spanish-surnamed panelists to the Rand Corporation, by the Civil Rights Commission, and by Dr. Laosa formed the basis for the evaluation both of the national battery and of site-specific instruments. The search for site-specific measures was made difficult by the fact that only experimental versions of tests exist in Navajo, and the fact that very few Spanish-language tests exist for preschool children.

The Head Start Collection report on tests for Spanish-speaking children (Rosen, 1971) yielded only two tests in Spanish for preschool children, neither of which was considered appropriate for PDC goals; one, because it required two days of testing and was only in experimental form (Inter-American Series--Manuel, 1966), the other, because it was primarily a cognitive battery (Walker Readiness Test for Disadvantaged Preschool Children--Walker, no date). The Thorpe Developmental Inventory (Thorpe, 1972), a developmental screening procedure, was also examined, but the McCarthy Scales of Children's Abilities had several advantages compared with the Thorpe even though the Thorpe has been used with Spanish-speaking children in California. The Van Alstyne Picture Vocabulary Test reviewed by Cruz (Raizen and Bobrow, 1974b), appeared to be "reliable and valid for the measurement of mental ability of culturally disadvantaged Mexican-American 6-year-olds," but the emphasis of PDC on social competence and not simply on cognitive ability led to the decision not to recommend that instrument, even if it were appropriate for 4-year-olds (see ratings in Table 5).

Recommended Measures

One instrument appeared to be relevant to the panelists' recommendation to measure language dominance (the Bilingual Syntax Measure--Burt, Dulay, and Hernandez, 1975) the BSM is designed

Table 5

Ratings of Measures for Spanish-Speaking Children on Criteria for Measures Selection

Measure	Practical Considerations					Psychometric Qualities					Relevance to PDC				Past Use
	Available for use by fall 1975	Appropriate for trained paraprofessionals	Test format appropriate for PDC age group	Scoring procedures appropriate for data processing	Reasonable testing time for young children	Adequate construct and/or predictive validity	Adequate test stability and internal consistency	Culture and/or SES fair	Representativeness of standardization sample	Low correlation with index of general information	Spans age range of PDC population (4-8)	Spanish adaptation available	Relevant to PDC Bilingual/Bicultural goals	Likely to demonstrate program effects	Used in previous national evaluation or large-scale studies
*Walker Readiness Test for Disadvantaged Preschool Children		X	X									X			
*Inter-American Series		X	X									X			
Thorpe Developmental Inventory		X	X	X	X			X	X			X	X		X
*Van Alstyne Picture Vocabulary Test															
Bilingual Syntax Measure	X	X	X	X	X						X	X	X		

*Technical manual has not yet been reviewed.

to measure children's oral proficiency in English and/or Spanish grammatical structures by using natural speech as a basis for making judgments. It encourages children to express their thoughts and opinions freely, and is designed primarily for children who are between the ages of 4 and 9. To elicit natural speech, simple questions are used with cartoon-type pictures to provide the framework for a conversation with the child. An analysis of the child's speech in answering these questions yields a numerical indicator and a qualitative description of the child's structural language proficiency.

It is recommended that both Spanish and English forms of the Bilingual Syntax Measure be administered to all children in the bilingual/bicultural sites, giving one form first to half of the children, and the other form first to the other half of the children, to obtain a measure of language dominance. It can also be used to measure the degree of maintenance or loss of certain basic Spanish grammatical structures in children who acquired Spanish as a first language.

Bilingual/bicultural education should also have a direct impact on classroom interactions experienced by teachers and children. Luis Laosa (personal communication, 1975) suggested that several categories could be added to the classroom observation system. The following dimensions will be developed into observational categories or ratings:

- Teacher's use of her aide, i.e., the role of bilingual aides;
- Relative frequency with which the two languages are actually used in the classroom and which language is spoken during different social interactions;
- Type of bilingual education model being used;
- Kinds of behaviors teachers reward and relative amount of teacher attention given to Spanish- and English-speaking children.

Adapting the Basic Battery for Bilingual/Bicultural Sites

All of the instruments recommended for the basic battery are currently being translated or adapted into Spanish. As indicated above, each of the measures in the basic battery will be carefully examined by the indigenous examiners to determine if the Spanish translations of the instructions are clear and if the items are appropriate to their region.

The observation instrument will include items on language pattern usage at school, and specific teacher-child interactions and teacher-aide interactions pertinent to language usage will be included. Again, the indigenous examiners are expected to be utilized in the interpretation of scores in the light of specific cultural expectations.

Measures for Site-Specific Batteries

The original design for the PDC evaluation recognized the fact that there would likely be many differences among programs implemented in 15 sites across the country. To allow for differences in site characteristics and in goals and objectives of individual programs, the evaluation design specified that, in addition to a basic battery to assess impact across all sites, site-specific batteries be employed to assess impact within programs. This section of the report presents the recommendations for measures to be used in site-specific batteries after reviewing the various issues that were taken into consideration in developing these recommendations.

Issues in the Development of Site-Specific Batteries

There are several issues related to the use of site-specific batteries in the evaluation of a national demonstration project. These are briefly discussed before presenting the recommendations for the batteries.

First, the nature of the basic experimental design and the analyses that accompany that design is involved. If substantial differences among sites exist in the outcome measures, the covariates, or in the correlations between covariates and outcome measures, then generalizations about the "national" program may be unwarranted. When this is the case, there is the option of treating the study as multiple site-specific studies or limiting the analyses to within-site analyses. This, of course, does not require different sets of measures for each site, but to the extent that site differences demand site-specific analyses, the study would be further strengthened by having measures that are particularly suited to the individual sites.

Built into the plan for the PDC evaluation is the pilot year of testing (1975-76). Given this flexibility it may be possible to consider using some site-specific measures not only to determine whether they are adequate for that particular site, but also to see whether they would be suitable measures for use in a number of sites or for inclusion in the basic battery used across all sites. This is not a sufficient justification for expending resources to develop site-specific measures, but is a use that should be kept in mind when site-specific measures are being recommended.

National policy issues also seem related to the use of site-specific batteries. On the national level there presumably are

concerns about the feasibility of a program with the features of Developmental Continuity. One concern might be whether the same (or essentially the same) program can be successfully implemented across the country regardless of local conditions or whether the program must be carefully tailored to individual sites' needs in order to be successful. Local successes may very well be obscured if the evaluation is limited to a national battery designed for the "average" site but less suited to any particular site.

Program Goals

A basic rationale for including in the evaluation measures that are unique to some sites is to enable the assessment of goals that are unique to those sites. The program responses to the goals questionnaire, which have been discussed in connection with the basic battery, have provided some information about the uniqueness of local goals. The information is somewhat difficult to interpret, however. Of the 25 goals for child competence and for parents and staff, there were very few on which a majority of the sites agreed. This is in part due to the questionnaire's response format which asked programs to indicate the five goals for children and the five for parents and staff considered "most important" for their program. But since the goals listed on the questionnaire were selected because they seemed to relate to PDC goals or to expert opinion regarding desirable program outcomes, even the ones not selected as "most important" might be expected to be valued by the programs. Thus, in developing the basic battery to be used across all sites, unanimity was not required when judging a goal as relevant for the national battery (see Figures 4 and 5). When responses to the goals questionnaire were examined with the site-specific battery in mind, there were only a few goals that were found to be uniquely held by a small number of sites. The following child goals, for example, were identified as "most important" by only one or two programs (of the 11 responding):

- Gross motor skills
- Perceptual skills
- Perceptual motor skills
- Auditory skills
- Flexibility in the application of information processing strategies
- School readiness
- Aural comprehension
- Productive (language) competence

Half of these are in the psychomotor area and half in the cognitive and language domain. The psychomotor goals, except for auditory skills, are being assessed as part of the basic battery because of the desire to have a battery that was responsive to the total range of social competence. Of the four cognitive and language

goals, aural comprehension is related to the verbal memory task recommended as part of the basic battery (see Table 4). Thus, on the basis of responses to the goals questionnaire, there are three areas identified by individual programs as being important that would be candidates for assessment with site-specific measures:

- Flexibility in the application of information processing strategies
- School readiness
- Productive language competence

The selection of measures related to these goals will be discussed below.

There were also several goals for parents and staff that were uniquely held by a few programs. However, all these goals are considered important in terms of national PDC guidelines, so it is recommended that they be assessed across all sites even if not considered particularly important by the local program (see Figure 3).

Reduced Importance of Site-Specific Batteries

Five factors have served to diminish the importance of having a complex set of site-specific measures. First, it has been feasible to recommend a fairly comprehensive basic battery that appears to address the concerns of a sizeable proportion of the sites. Second, the information on site-specific goals obtained from the goals questionnaire suggests that there are relatively few goals that are unique to one or two programs. Third, the bilingual/bicultural measures recommended in the preceding section of this report successfully address a large portion of the site-specific objectives and concerns of the bilingual/bicultural demonstration programs. A fourth factor is that at the elementary level, where the effects of continuity are first manifested, all school systems conduct their own testing programs. It is possible (though not always the case) that these tests will be relevant to PDC goals and that data from them can be incorporated into site-specific evaluations. Finally, one of the purposes of site-specific measures can be accomplished in the analysis and interpretation of findings from the basic battery. Local programs can be given the opportunity to indicate particular tests, subtests, or even individual items which they judge as assessing goals that are particularly important to their program. These can then be treated in the analysis as "site-specific measures" for the purpose of making judgments on the effectiveness of programs in terms of their own goals. This would be a valuable procedure for another reason. Decisions about measures to be used have been based on judgments of the relevance of the measures to various goal statements (as, for example, listed in the goals questionnaire). Since the goal statements are not operationalized, they are open

to different interpretations and even though two programs may appear to agree on a goal, they may in fact have different objectives. When program staff have an opportunity to examine specific items and tests they will be in a better position to judge the relevance of the measure to their program.

Although for the purposes of this report recommendations for site-specific measures seem less important than previously, it should be made clear that this conclusion is not intended to diminish the importance of differences in site characteristics and goals. The need to provide an evaluation process that is responsive to local concerns is still paramount and will be addressed in the ways discussed above.

Recommended Measures for Site-Specific Goals

In light of the above considerations, a comprehensive set of measures to assess potential site-specific concerns does not seem warranted. Rather, recommendations will be made for measurement in key areas; the final selection will, however, be at the option of the sites. The areas are:

- School readiness
- Productive language competence
- Flexibility in the application of information processing strategies

School readiness. The initial step in selecting an instrument to measure "readiness" was to define the behaviors or abilities which constitute "school readiness." Anderson and Messick (1974), preschool curriculum literature, and child development experts were consulted in arriving at the following definition:

> School readiness typically encompasses the child's age-appropriate acquisition of fine visual discrimination skills; beginning recognition of numbers and letters; and a basic understanding of quantitative concepts. In addition, readiness implies the child possesses a reasonable degree of knowledge in areas important to functioning in and out of school. This might include a basic understanding of physical laws, rules of health and safety, the physical and social environment, and practical arts.

In a review of other Head Start evaluations and the battery being considered for the National Day Care study several preschool-kindergarten instruments were identified which could be further considered. Of these, the Peabody Picture Vocabulary Test, Preschool Inventory, and Denver Developmental Screening Test received consideration for possible use in site-specific batteries. Although its use is restricted to small scale research and a standardization study, the CIRCUS (Anderson, et

al., 1974), with high recommendations from OCD, also became a potential candidate. When compared in terms of content, scope and psychometric characteristics with the above preschool measures, the CIRCUS is seen as similar, but in some respects more sophisticated. Although the standardization sample was heavily biased toward white, middle-SES children, the use of this instrument in the PDC evaluation may generate valuable information concerning its appropriateness for lower-SES, black children. Since this information might also contribute to future Head Start evaluations, certain CIRCUS subtests are recommended for use in site-specific batteries. A brief description of the instrument is provided below and more detailed description of its psychometric qualities is provided in the appendix.

The CIRCUS is an assessment battery developed by Educational Testing Service (Anderson, et al., 1974) for preprimary children. The battery includes 16 diagnostic measures designed to assess specific cognitive, language and perceptual motor characteristics of preschool and kindergarten children. The authors used the theme of a circus to represent their basic pencil and paper testing format. This may appeal to the interest of the preprimary child, thus assisting in maintaining his/her attention. With the exception of two subtests all can be group-administered by a teacher or trained paraprofessional. Currently, a Spanish version of the CIRCUS is under development by ETS.

Although the authors describe the CIRCUS as a "comprehensive, flexible program of assessment services for preprimary children," testing constraints and the PDC selection criteria led to selecting only those subtests which assess school readiness rather than recommending the complete battery. This is possible partly because the authors designed their instrument to allow the separate use of various subtests to measure specific areas. The subtests recommended for assessment of school readiness are listed and described below. A fuller account of their content and psychometric features may be found in the appendix.

- How Much and How Many. This subtest was developed to assess the child's understanding of quantitative concepts. The child is first asked to select from among three choices the picture which correctly depicts the number of objects that a numeral represents. The second component asks the child to choose the longer, smallest, first or bottom object in order to assess his/her comprehension of quantitative vocabulary.

- Finding Letters and Numbers. This subtest asks the child to select from among three choices the letter or number named by the tester.

- How Words Work. This subtest focuses on the structure (syntax or word order) of language and the function words (prepositions, articles, pronouns, conjunctions) which a sentence may contain in order to assess the child's receptive vocabulary. The child is asked to choose between two pictures the one that the tester has described in a sentence.

- Do You Know...? This subtest contains information items on health and safety, physical and social environment, consumer behavior or other factors relevant to the child's functioning in and out of school. The tester requests the child to choose from among three options that picture which best applies to a given question or statement. For example, the child may be asked, "Which one of these is SAFE to play with?" The child then chooses the appropriate response from pictures of a book of matches, a knife and a spoon.

While the administration of these subtests in their complete form might be desirable for evaluation purposes, the total time involved (approximately 60 minutes) would cause the total testing time to become excessive at sites where these objectives are to be assessed. Since PDC sites receiving this battery are implementing curricula which emphasize different aspects of school readiness, each site may elect only one or two subtests.

Productive language competence. Language experts point out the dangers of attempting to assess "natural" language in a standardized setting. Nevertheless, the importance of the attempt is not denied. One widely used measure (the ITPA Verbal Expression subtest--Kirk, McCarthy, and Kirk, 1968; Paraskevopoulos and Kirk, 1969) does not appear to assess the type of production described in the goals questionnaire: "competence in the narrative use of language demonstrated by creatively telling a story about a given event or experience." Furthermore, the ITPA has received serious criticism from a number of quarters (e.g., Carroll, 1972; Severson and Guest, 1970).

Competence in narrative language has been assessed in a number of Follow Through evaluations using procedures developed by some of the Follow Through sponsors (e.g., Abrams, Rhodes, and Tanaka, 1973; Gould, 1974; Love, 1975), but they have not been adapted for preschool-aged children.

The Say and Tell subtest of CIRCUS has a part that assesses "narrative use of language" by asking the child to make up a story about a drawing of a busy circus ring. The story is recorded verbatim and scored according to the number of words, number of different words, use of labels, verbs, modifiers, syntax, sequence, plot extension, organization, feeling, rhythm

and cadence, comparison, character extension and spatial words. There appears to be some question as to the suitability of either the story elicitation procedure or the scoring criteria since 7 of the 12 criteria were seldom seen in the preschool and kindergarten stories (fewer than 20% of the stories contained these elements). Furthermore, there appeared to be little difference between preschool and kindergarten levels in the percent of stories containing the elements. This was also true for three summary measures: total words (mean of 41.7 for preschool, 43.1 for K); number of different words (26.2 for preschool, 27.1 for K); and narration quality (4.12 for preschool, 4.11 for K). The small difference between children one year apart suggests the measure may not be particularly sensitive to program effects. Experience in using the same CIRCUS picture to elicit written stories from older children in Follow Through suggests that the picture may be too "busy" and not appropriate for eliciting organized, sequenced stories (Abrams, Rhodes, and Tanaka, 1973).

Since productive language is an important goal for at least one PDC site (and it may be for sites that did not respond to the goals questionnaire) and since several Follow Through sponsors have completed considerable developmental work of productive language measures for the elementary grades, it is recommended that Say and Tell be listed as an option for site-specific batteries so that additional psychometric data can be obtained during the pilot phase. In addition, however, it is recommended that an adaptation of the Follow Through procedures be developed for Head Start-age children and made available on an experimental basis this fall to sites that are particularly concerned with this goal. By carefully considering the story elicitation technique (e.g., real objects might be given to children to play with before asking them to narrate a story) and scoring criteria appropriate for Head Start-age children, it should be possible to have a testing procedure available by fall which will address this site-specific concern.

Flexibility in the application of information processing strategies. Anderson and Messick defined this flexibility as follows: "The child recognizes that there are different approaches to exploring the environment and to obtaining and processing information from it, he recognizes that these approaches are differentially effective in different situations, and he applies these approaches flexibly and appropriately...without being locked into habitual modes of perceiving and thinking" (Anderson and Messick, 1974, p. 291).

The Rand report discussed several constructs related to this competency by considering flexibility in response to both nonpersonal and interpersonal stimuli. Several measures were recommended by Rand, but most of them were developed as experimental techniques to measure components of "flexibility." A disadvantage shared by all the measures is the artificiality of the structured testing situation. In the Concept-Switching

Task of Zigler and deLabry (1962), for example, the child is asked to sort and resort a set of concept cards. In addition to representing only a narrow aspect of flexibility, there is some question as to whether the decentering required by such a task makes it suitable for preoperational children. In another area, presenting children with situations where multiple solutions are possible (e.g., the "unusual uses" type of test), Rand found no measures to recommend.

If flexibility of response to interpersonal stimuli is considered, the PIPS test, which is recommended as a measure of social problem-solving, may very well be relevant to the "flexibility" goals of some programs. The classroom observation system also includes behavior categories that address aspects of flexibility: Category 6 ("uses peer as resource") codes whether the child asks for help before or after attempting to solve a problem.

Although these few measures may not satisfy the needs of programs with a strong interest in assessing this competency, given the measurement problems in this area and the constraints on test development efforts, it is recommended that no additional measures be selected for the assessment of "flexibility in the application of information processing strategies."

Auditory discrimination. As mentioned in the section on psychomotor measures, no auditory descrimination or visual perception measure is recommended for the basic battery. Thus, to accommodate sites that might be interested in this particular psychomotor area, the Wepman Auditory Discrimination Task will be available as a site-specific option.

Procedures for Implementing Site-Specific Batteries

Since site-specific measures are, by definition, necessary only if sites desire to be assessed with those measures, a procedure has been developed for obtaining input from sites regarding the proposed measures. A description of each instrument and an outline of procedures will be prepared by July 15. These will be sent to all PDC programs with a letter describing the purpose of the site-specific batteries, and a request for a response by August 1 if the site wishes one or more of these measures to be included in their battery. This amount of lead time is needed to prepare the necessary test booklets and materials and to make arrangements for tester training. Sites will have to be somewhat limited in the number of measures they can ask to be evaluated with since there are limitations in the total time available for testing.

IV

SUMMARY OF RECOMMENDED MEASURES

Social-Emotional Measures

Child Interview*
Preschool Interpersonal Problem-Solving Test
Stephens-Delys Reinforcement Contingency Interview
Pupil Observation Checklist
Child Rating Scale*
PDC Classroom Observation System*

Psychomotor Development, Health and Nutrition Measures

Block Building (MSCA)
Draw-A-Child (MSCA)
Conceptual Grouping (MSCA)
Leg Coordination (MSCA)
Arm Coordination (MSCA)
Block Design (WISC)
Block Design (WPPSI)
Health and Nutrition Records

Cognitive and Language Measures

Block Design (WISC)
Block Design (WPPSI)
Conceptual Grouping (MSCA)
Verbal Memory (MSCA)
Verbal Fluency (MSCA)
Opposite Analogies (MSCA)

Teacher and Parent Measures

Teacher Survey*
Parent Survey*

Bilingual/Bicultural Measures

Bilingual Syntax Measure
PDC Classroom Observation System*

Measures for Site-Specific Batteries

How Much and How Many (CIRCUS)
Finding Letters and Numbers (CIRCUS)
How Words Work (CIRCUS)
Do You Know (CIRCUS)
Say and Tell (CIRCUS)
Wepman Auditory Discrimination Test

---

*To be developed.

V

REFERENCES

Abrams, A., Rhodes, A., & Tanaka, M. Assessing written language production in third grade children. San Francisco: Far West Laboratory for Educational Research and Development, 1973.

Anderson, S. B., Beaton, A., Emmerich, W., & Messick, S. Disadvantaged children and their first school experiences, ETS-OEO Longitudinal Study: Theoretical considerations and measurement strategies. Princeton, N.J.: Educational Testing Service, 1968.

Anderson, S. B., Bogatz, G. A., Draper, T. W., Jungeblut, A., Sidwell, G., Ward, W. C., & Yates, A. CIRCUS manual and technical report (preliminary version). Princeton, N.J.: Educational Testing Service, 1974.

Anderson, S., & Messick, S. Social competency in young children. Developmental Psychology, 1974, 10, 282-293.

Birch, H. G., & Gussow, J. D. Disadvantaged children: Health, nutrition, and school failure. New York: Harcourt, Brace, & World, 1970.

Bogatz, G. A., & Ball, S. The second year of Sesame Street: A continuing evaluation (Vol. 1). Princeton, N.J.: Educational Testing Service, 1971.

Boger, R., & Knight, S. Socio-emotional task force (Final Report). Washington, D.C.: Office of Economic Opportunity, U.S. Department of Health, Education, and Welfare, 1969.

Borke, H. Chandler and Greenspan's "Ersatz Egocentrism": A rejoinder. Developmental Psychology, 1972, 7, 107-109.

Bronson, M. Manual for executive and social skill profile for preschool children. Cambridge, Mass.: Harvard University, Graduate School of Education, 1973.

Bronson, M. Executive competence in preschool children. Paper presented at the meeting of the American Educational Research Association, Washington, D.C., April 1975.

Brown, B. The Brown-IDS self-concept referents test. New York: New York University, Institute for Developmental Studies, 1966.

Buros, O. K. (Ed.). The sixth mental measurements yearbook. Highland Park, N.J.: Gryphon Press, 1965.

Buros, O. K. (Ed.). The seventh mental measurements yearbook (2 vols.). Highland Park, N.J.: Gryphon Press, 1972.

Burt, M., Dulay, H., & Hernandez, E. Bilingual syntax measure. New York: Harcourt, Brace, Jovanovich, 1975.

Butler, A. L., Gotts, E. E., Quisenberry, N. L., & Thompson, R. P. Literature search and development of an evaluation system in early childhood education. Bloomington, Ind.: Indiana University, 1971.

Carroll, J. B. Review of the Illinois Test of Psycholinguistic Abilities. In O. K. Buros (Ed.), The seventh mental measurement yearbook (2 vols.). Highland Park, N.J.: Gryphon Press, 1972.

Chance, J. E. Academic correlates and maternal antecedents of children's belief in external or internal control of reinforcement. In J. B. Rotter, J. E. Chance & E. J. Puares (Eds.), Applications of a social learning theory of personality. New York: Holt, Rinehart, & Winston, 1972.

Chandler, M. J., & Greenspan, S. Ersatz egocentrism: A reply to H. Borke. Developmental Psychology, 1972, 7, 104-106.

Chissom, B. S. Review of the Test of Visual-Motor Integration. In O. K. Buros (Ed.), The seventh mental measurements yearbook (2 vols.). Highland Park, N.J.: Gryphon Press, 1972.

Cline, M. G. Education as experimentation: Evaluation of the Follow Through planned variation model (Vol. 1-B, Final Report). Cambridge, Mass.: Abt Associates Inc., 1974.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. Equality of educational opportunity. Washington, D.C.: U.S. Government Printing Office, 1966. (Superintendent of Documents, Catalog No. FS 5.238.38001)

Coller, A. R. The assessment of 'self-concept' in early childhood education. Urbana, Ill.: ERIC Clearinghouse on Early Childhood Education, 1971. (ERIC Document Reproduction Service No. ED 050 822)

Coughlan, R. J. The school survey. Chicago: University of Chicago, Industrial Relations Center, 1966.

Coughlan, R. J., & Cooke, R. A. Work attitudes. In H. J. Walberg (Ed.), Evaluating educational performance: A sourcebook of methods, instruments, and examples. Berkeley, Calif.: McCutchan, 1974.

Crandall, V. C., Katovsky, W., & Crandall, V. J. Children's belief in their own control of reinforcements in intellectual-academic achievement situations. Child Development, 1965, 36, 91-109.

Cronbach, L. J., Gleser, G. C., Nauda, H., & Rajaratnam, N. The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: John Wiley, 1972.

Crowne, D. D., & Stephens, M. W. Self-acceptance and self evaluative behavior: A critique of methodology. Psychological Bulletin, 1961, 58, 104-121.

Deloria, D., Love, J., Goedinghaus, L., Gordon, S., Hanvey, R., Hockman, E., Platt, J., & Nauta, M. The National Home Start Evaluation: Summative evaluation results (Interim Report IV). Ypsilanti, Mich.: High/Scope Educational Research Foundation, 1974.

DiCarlo, L. M. Review of the Wepman Auditory Discrimination Test. In O. K. Buros (Ed.), The sixth mental measurments yearbook. Highland Park, N. J.: Gryphon Press, 1965.

Emrick, J. A., Sorensen, P. H., & Stearns, M. S. Evaluation of the National Follow Through Program: Results of the winter 1971 pretest for the spring 1972 Follow Through non-cognitive test battery. Menlo Park, Calif.: Stanford Research Institute, 1972.

Feshbach, N. D., & Feshbach, S. The relationship between empathy and aggression in two age groups. Developmental Psychology, 1969, 1, 102-107.

Fitts, W. H. Tennessee self-concept scale. Nashville, Tenn.: Counselor Recordings and Tests, 1964.

Fountain, P. What teaching does to teachers: The teacher as worker. Unpublished doctoral dissertation, Yale University, 1975.

Gould, D. Coding manual for the use of ACOST (Analysis of Child Oral Story Telling). New York: Bank Street College of Education, 1974.

Herbert, J. Direct observation as a research technique. Psychology in the Schools, 1970, 7, 127-138.

Hess, R. D., Shipman, V. C., Brophy, J. E., & Bear, R. M. The cognitive environments of urban preschool children. Chicago: University of Chicago, 1969.

Higham, J. Send these to me: Jews and other immigrants in urban America. New York: Atheneum, 1975.

Hoepfner, R., Stern, C., & Nummedal, S. G. CSE-ECRC Preschool/Kindergarten test evaluations. Los Angeles: Center for the Study of Evaluation, 1971.

Jamison, C. B. Review of the Purdue Perceptual Motor Survey. In O. K. Buros (Ed.), The seventh mental measurements yearbook (2 vols.). Highland Park, N.J.: Gryphon Press, 1972.

Johnson, O. G., & Bommarito, J. W. Tests and measurements in child development: A handbook. San Francisco: Jossey-Bass, 1971.

Kaufman, A. S. Comparison of the WPPSI, Stanford-Binet, and McCarthy Scales as predictors of first-grade achievement. Perceptual and Motor Skills, 1973, 36, 67-73.

Kaufman, A. S. Analysis of the McCarthy Scales in terms of Guilford's structure of intellect model. Perceptual and Motor Skills, 1973, 36, 967-976.

Kaufman, A. S., & Hollenbeck, G. P. Factor analysis of the standardization of the McCarthy Scales. Journal of Clinical Psychology, 1973, 29, 358-362.

Kaufman, A. S., & Kaufman, N. L. Black-white differences at ages 2-1/2 to 8-1/2 on the McCarthy Scales of Children's Abilities. Journal of School Psychology, 1973, 11, 196-206.

Kirk, S. A., McCarthy, J. J., & Kirk, W. D. Illinois Test of Psycholinguistic Abilities: Examiner's manual (Rev. ed.). Urbana, Ill.: University of Illinois Press, 1968.

Klein, S. W. Making or breaking it: The teacher's role in model curriculum implementation. Young Children, 1973, 28, 359-366.

Laosa, L. What languages do bilingual children use with whom? Research evidence and implications for education. Paper presented at the 1975 Convention of the Association of Teachers of English to Speakers of Other Languages (TESOL), Los Angeles, March 1975.

Littell, W. M. The Wechsler Intelligence Scale for Children: Review of a decade of research. Psychological Bulletin, 1960, 57, 132-156.

Long, B. H., & Henderson, E. H. Social schemata of school beginners: Some demographic correlates. Merrill-Palmer Quarterly, 1970, 16, 305-324.

Love, J. M. The High/Scope Productive Language Arts Test in the Cognitive Curriculum: Effects of Follow Through on the written language of second and third grade children. Ypsilanti, Mich.: High/Scope Educational Research Foundation, 1975.

Manuel, H. T. Inter-American series. Austin, Tex.: Guidance Testing Associates, 1966.

McCarthy, D. McCarthy Scales of Children's Abilities: Manual. New York: Psychological Corporation, 1972.

Nadler, D., Canman, C., Jerkins, G. D., & Laecter, E. E. The Michigan assessment package. Ann Arbor, Mich.: University of Michigan, Institute for Social Research, 1974.

Nowicki, S., Jr., & Roundtree, J. Correlates of locus of control in secondary school age students. Developmental Psychology, 1971, 4, 477-478.

Nowicki, S., Jr., & Strickland, B. R. A locus of control scale for children. Journal of Consulting and Clinical Psychology, 1973, 40, 148-154.

Ogilvie, D., & Shapiro, B. Manual for assessing social abilities of one to six year old children. In B. L. White & J. C. Watts (Eds.), Experience and environment: Major influences on the development of the young child. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Paraskevopoulos, J. N., & Kirk, S. A. The development and psychometric characteristics of the revised Illinois Test of Psycholinguistic Abilities. Urbana, Ill.: University of Illinois Press, 1969.

Piers, E., & Harris, D. Manual for the Piers-Harris Self Concept Scale. Nashville, Tenn.: Counselor Recordings and Tests, 1969.

Prescott, E., & Jones, E. Day care as a child-rearing environment. Washington, D.C.: National Association of the Education of Young Children, 1972.

Raizen, S., & Bobrow, S. B. Design for a national evaluation of social competence in Head Start children. Santa Monica, Calif.: Rand Corporation, 1974. (a)

Raizen, S., & Bobrow, S. B. Appendices to design for a national evaluation of social competence in Head Start children. Santa Monica, Calif.: Rand Corporation, 1974. (b)

Rayder, N. F., & Bödy, B. The Educational Forces Inventory: A new technique for measuring influences on the classroom. Unpublished manuscript, 1975. (Available from Far West Laboratory for Educational Research and Development, 1855 Folsom Street, San Francisco, California 94103). (a)

Rayder, N. F., & Bödy, B. The Educational Forces Inventory: Psychometric properties. Unpublished manuscript, 1975. (Available from Far West Laboratory for Educational Research and Development, 1855 Folsom Street, San Francisco, California 94103). (b)

Rentfrow, R. K., Durning, K., Conrad, E., & Goldupp, O. Intensive evaluation of Head Start implementation in the Tucson Early Education Model. Tucson, Ariz.: University of Arizona, College of Education, 1972.

Ringelheim, D., Bialer, I., & Morissey, H. Final report (Report No. 6-2685). Washington, D.C.: U.S. Department of Health, Education, and Welfare, 1970.

Rivlin, A. M., & Timpane, P. M. Planned variation in education: Should we give up or try harder? Washington, D.C.: Brookings Institution, 1975.

Robinson, J. P., & Shaver, P. R. Measures of social psychological attitudes (Rev. ed.). Ann Arbor, Mich.: University of Michigan, Institute for Social Research, 1974.

Rosen, P. Tests for Spanish-speaking children: An annotated bibliography. Princeton, N.J.: Educational Testing Service, 1971.

Ross, S., & Zimilies, H. The Differentiated Child Behavior Observational System (DCB). In M. C. Wang (Ed.), The use of direct observation to study instructional-learning behaviors in school settings. Pittsburgh, Pa.: University of Pittsburg, Learning Research and Development Center, 1974.

Rotter, J. B. Generalized expectancies for internal versus external control of reinforcement. Psychological Monographs, 1966, 80(1, Whole No. 609).

Schmuck, R. A., & Miles, M. B. (Eds.). Organization development in schools. Palo Alto, Calif.: National Press, 1971.

Schmuck, R. A., Runkel, P. J., Saturen, S. L., Martell, R. T., & Derr, C. B. Handbook of organization development in schools. Palo Alto, Calif.: National Press, 1972.

Severson, R. A., & Guest, K. E. Toward the standardized assessment of the language of disadvantaged children. In F. Williams (Ed.), Language and poverty: Perspectives on a theme. Chicago: Markham, 1970.

Shipman, V. C., & Gilbert, L. E. Brown IDS Self-Concept Referents Test (Tech. Rep. 2). In V. C. Shipman (Ed.), Disadvantaged children and their first school experiences: Technical report series. Princeton, N.J.: Educational Testing Service, 1972.

Shure, M. B., & Spivack, G. The PIPS test manual. Philadelphia, Pa.: Hahnemann Medical College, 1974.

Slobin, D. I. (Ed.). A field manual for cross-cultural study of communicative competence (2nd draft). Berkeley, Calif.: University of California, 1967.

Smith, M. S. Some short term effects of Project Head Start: A preliminary report on the second year of planned variation--1970-71. Cambridge, Mass.: Huron Institute, 1973.

Sorensen, P. H., & Madow, W. Longitudinal evaluation of selected features of the National Follow Through Program. Menlo Park, Calif.: Stanford Research Institute, 1971.

Spivak, G., & Shure, M. Social adjustment of young children. San Francisco, Calif.: Jossey-Bass, 1974.

Stanford Research Institute. Implementation of planned variation in Head Start. Menlo Park, Calif.: Author, 1971.

Stanford Research Institute. Results of winter 1971 pretest for the spring 1972 Follow Through non-cognitive test battery. Menlo Park, Calif.: Author, 1972.

Stephens, M. W., & Delys, P. A locus of control measure for preschool children. Developmental Psychology, 1973, 9, 55-65. (a)

Stephens, M. W., & Delys, P. External control expectancies among disadvantaged children at preschool age. Child Development, 1973, 44, 670-674. (b)

Thorpe, H. S. Developmental appraisal of the preschool child ages three to six: The Thorpe Developmental Inventory. Davis, Calif.: University of California, 1972.

U.S. Commission on Civil Rights. A better chance to learn: Bilingual-bicultural education. Washington, D.C.: Author, 1975. (Clearinghouse Publication No. 51)

Wacker, R. F. Race and ethnicity in American social science, 1900-1950. Unpublished doctoral dissertation, University of Michigan, 1975.

Walberg, H. J. (Ed.). Evaluating educational performance. Berkeley, Calif.: McCutchan, 1974.

Walker, D. K. Socioemotional measures for preschool and kindergarten children. San Francisco, Calif.: Jossey-Bass, 1973.

Walker, D. K., Bane, M. J., & Bryk, A. The quality of the Head Start planned variation data. Cambridge, Mass.: Huron Institute, 1973.

Walker, W. Walker Readiness Test for Disadvantaged Preschool Children. Kansas City, Mo.: Regional Research Program, no date. (ERIC Document Reproduction Service No. ED 037 253)

Wechsler, D. Wechsler Intelligence Scale for Children: Manual. New York: Psychological Corporation, 1949.

Wechsler, D. Wechsler Preschool and Primary Scale of Intelligence: Manual. New York: Psychological Corporation, 1967.

Werner, E. E. Review of the Gesell Developmental Schedules. In O. K. Buros (Ed.), The sixth mental measurements yearbook. Highland Park, N.J.: Gryphon Press, 1965.

White, B. L., & Watts, J. C. (Eds.). Experience and environment: Major influences on the development of the young child. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

White, B., et al. The preschool project: Experience and the development of human competence in the first six years of life. Cambridge, Mass.: Harvard University, Center for Research and Development on Educational Differences, 1969.

Woo-Sam, J. M., & Zimmerman, I. L. Research with the Wechsler Preschool and Primary Scale of Intelligence (WPPSI): The first five years. School Psychology Monograph, 1973, 1(1), 25-50.

Wright, H. F. Observational child study. In P. H. Mussen (Ed.), Handbook of research methods in child development. New York: Wiley, 1960.

Zigler, E., & deLabry, J. Concept-switching in middle-class, lower-class, and retarded children. Journal of Abnormal and Social Psychology, 1962, 65, 267-273.

Zimmerman, I. L., & Woo-Sam, J. Research with the Wechsler Intelligence Scale for Children: 1960-1970. Journal of Clinical Psychology Monograph, Supplement 33, 1972.

APPENDIX :

INSTRUMENT REVIEWS

Page

Social-Emotional Measures

Preschool Interpersonal Problem-Solving Test. 87
Stephens-Delys Reinforcement Contingency Interview. . 89
Pupil Observation Checklist 91

Psychomotor Development

McCarthy Scales of Children's Abilities 92
Block Design (WISC) 93
Block Design (WPPSI). 94

Cognitive and Language Measures

McCarthy Scales 92
Block Design (WISC) 93
Block Design (WPPSI). 94

Bilingual/Bicultural Measure

Bilingual Syntax Measure. 95

Site-Specific Measures

How Much and How Many 96
Finding Letters and Numbers 97
How Words Work. 98
Do You Know...? 99
Say and Tell. 100
Wepman Auditory Discrimination Test 101

Preschool Interpersonal Problem-Solving Test

PUBLISHER OR AUTHOR: Shure and Spivak
ADDRESS: Hahnemann Medical College and Hospital

CONTENT: Social-emotional: social problem-solving

DESCRIPTION: The PIPS attempts to assess the child's ability to name alternative solutions to two life-related types of problems: (a) ways for one child to obtain a toy from another and (b) ways to avert the mother's anger which could result from damage to property. Among inner city four year olds attending the Philadelphia Get Set day care program those judged as better adjusted by their teachers were able to conceptualize a greater number and a wider range of alternative solutions to real life problems than could their more poorly adjusted classmates irrespective of sex types. PIPS scores have also related to the child's general concern for the feelings of others (empathy), the degree to which he was liked by his/her peers, and to a lesser extent, the degree of general initiative and autonomy shown in the classroom.

AGE RANGE: 4 to 5 year olds

ADMINISTRATION: Test time: 20-25 minutes
Group or individual: Individual
Who administers: Trained tester
How administered: Interview
Response mode: Verbal
Comments: Multiethnic pictures are included to facilitate responses

VALIDITY: Validity for the PIPS is claimed by the authors because significant correlations were observed with teacher ratings of social adjustment. It appears that the test can discriminate between children who differ in the degree of behavioral adjustment exhibited in the classroom. There is a low but significant correlation between the Peabody picture vocabulary test (PPVT) and PIPS scores ($r = 38$, $p<.01$). No correlation was observed between degree of verbalization and PIPS.

RELIABILITY: Inter-rater reliability = .86-.89
Test-retest: 180 children were retested after a period of 3.5 months. Reliability coefficient = .59 according to the authors, this evidence indicates that PIPS test measures a property of thought that remains relatively stable for long periods of time.

NORMS: Over a four-year period, a total of 469 inner city four-year olds (220 boys and 249 girls) have been administered the PIPS between the months of November and January. Based on cumulative percentages it was possible to determine cut-off scores for differentiating between behavior categories of adjusted, impulsive and inhibited.

SOURCE IN WHICH DESCRIBED: Shure and Spivak. Problem-solving thinking and adjustment among disadvantaged. Child Development, 1971, 42, 179-180. Social Adjustment of Young Children, Spivak and Shure, 1974.

PERMISSION TO USE CAN BE OBTAINED FROM: Dr. Myrna Shure, Hahnemann Medical College and Hospital, Pittsburgh, Pennsylvania 19102.

Stephens-Delys Reinforcement Contingency Interview

PUBLISHER OR AUTHOR: Mark W. Stephens
ADDRESS: Apt. 647
1127 Lincoln Ave.
Evansville, Illinois 47114

CONTENT: Social-emotional: locus of control

DESCRIPTION: This measure taps a child's expectancy that his own behavior would change the probability that reinforcement might occur; that is, if he changes his behavior, will his teachers', parents' or friends' attitude or behavior change as a result. Forty questions have been used (but a shortened form of 20 questions is possible) posing reinforcing-type events, e.g., "What makes mothers smile?" Responses are coded "internal" if answered, "When I..." and "external" when answered, "When Daddy (or someone other than self)...". This test has been used with Head Start children as well as 575 second graders from Follow Through classes and middle-class schools. Subscales are possible for teachers, parents, and peers.

AGE RANGE: 4 years through 10 years

ADMINISTRATION: Test time: 10 to 25 minutes
Group or Individual: Individual
Who administers: Trained tester
How administered: Interview
Response mode: Verbal, free-response

VALIDITY: Discriminant: This test was related consistently with I.Q. tests of preschoolers.
Convergent: The evidence is inconsistent. Locus of control appears to be a multi-dimensional construct for children as well as adults.

RELIABILITY: Subscales intercorrelate .70.
Inter-rater reliability was .98.
Test-retest reliability after 4 months was .62.

NORMS: 41 Ss from Head Start classes were found to be significantly less internal than 45 middle-class children. Internality increases with age.
Disadvantaged Ss = 9.6. Advantaged Ss = 14.2.

SOURCE IN WHICH DESCRIBED: Robinson, J. P., and Shaver, P. R. Measures of social psychological attitudes. University of Michigan Institute for Social Research, Ann Arbor, 1974.

PERMISSION TO USE CAN BE OBTAINED FROM: Mark Stephens

Pupil Observation Checklist (POCL)

PUBLISHER OR AUTHOR: High/Scope Educational Research Foundation
ADDRESS: 600 N. River Street
Ypsilanti, MI 48197

CONTENT: Social-emotional: Test-taking behavior

DESCRIPTION: This is a rating scale of nine 7-point bipolar adjectives developed for the Home Start evaluation from a 25-item scale used in Follow Through. Two factors (labeled "Test Orientation" and "Sociability") were found consistently across several time points in the Home Start evaluation.

AGE RANGE: Used for 3 to 6 year old children; probably suitable for older children as well.

ADMINISTRATION: Test time: 5 minutes of tester's time per child
Group or individual: Child is not involved
Who administers: No training necessary
How administered: Rating scale
Response mode: NA

VALIDITY: No information

RELIABILITY: Internal consistency alpha above .90 for each scale; Test-retest correlation (7 month interval) found to depend on treatment group: Test orientation, above .60 for Home Start, .48 for control, and .49 for Head Start. Sociability, above .50 for Home Start and control group, .36 for the Head Start sample.

NORMS: Scale scores available for samples of Home Start, control and Head Start children at various time points, e.g., pretest scores available on a total of 552 children from age 3 to 5 1/2 at 6-month intervals.

SOURCE IN WHICH DESCRIBED: Deloria, D., et al., The National Home Start Evaluation; Interim Report IV: Summative Evaluation Results, Report to OCD, HEW, June 14, 1974.

PERMISSION TO USE CAN BE OBTAINED FROM: Research Department, High/Scope Foundation

McCarthy Scales of Children's Abilities

PUBLISHER OR AUTHOR: Dorothea McCarthy
ADDRESS: Psychological Corporation
304 East 45th Street
New York, New York 10017

CONTENT: Cognitive, Psychomotor

DESCRIPTIONS: Series of tasks tapping problem-solving, psychomotor, and conceptual abilities similar to the Wechsler scales, but with emphasis on age-related maturational indicators. Tests of general abilities, some of which may be subject to program effects.

AGE RANGE: 2-1/2 years through 8-1/2 years

ADMINISTRATION: Test time: 40 minutes total
Group or individual: Individual
Who administers: Trained tester
How administered: Testing
Response mode: Variable, depending on task--verbal and motor performance
Comments: Relatively few black-white differences on scales encourages use with disadvantaged

VALIDITY: Test is too new for major validation studies. Manual presents high positive correlation (.70-.85) of total score with WPPSI, Stanford-Binet, and moderate correlations with first grade achievement tests (.30-.49).

RELIABILITY: Intercorrelation of subtests and stability coefficients for individual subtests are not presented in the manual. Rentfrow, et. al. (1972) present re-test correlations and subtest-total correlations which are high and positive for most subtests chosen.

NORMS: Standardized on 1,032 children selected as representative of U.S. population (geography, urban-rural, ethnic background, father's occupation)

SOURCE IN WHICH DESCRIBED: McCarthy Scales of Children's Abilities Manual (McCarthy, Dorothea, 1970)

PERMISSION TO USE CAN BE OBTAINED FROM: Psychological Corporation

Wechsler Intelligence Scale for Children, Block Design (subtest)

PUBLISHER OR AUTHOR: David Weschler
ADDRESS: Psychological Corporation
304 East 45th Street
New York, New York 10017

CONTENT: Cognitive, Psychomotor

DESCRIPTION: Task consists of child's reproducing (constructing) designs with colored blocks (cubes) either modeled by the examiner or presented on a card. The measure taps problem-solving abilities, flexibility in response style, visual-motor organization and execution.

AGE RANGE: 5 years through 15 years

ADMINISTRATION: Test time: 10 minutes (timed items)
Group or individual: Individual
Who administers: Trained tester
How administered: Testing
Response mode: Manual performance, manipulation of objects

VALIDITY: Studies were not reported for this subtest alone. Concurrent validity of overall wise scores: correlates highly with Stanford-Binet (r=.80's) in many studies. Studies of predictive validity have not been reported.

RELIABILITY: Split half reliabilities at three age levels were presented in the manual, all in the .85 range. Block Design correlates in the .50-.65 range with the Full Scale I.Q. score at various ages.

NORMS: Standardized on a representative selection of U.S. White children (based on geography, urban-rural, and occupation). There were 2,200 children in the sample.

SOURCE IN WHICH DESCRIBED: Wechsler Intelligence Scale for Children Manual (Wechsler D., 1949)

PERMISSION TO USE CAN BE OBTAINED FROM: Psychological Corporation

Wechsler Preschool and Primary Scale of Intelligence, Block Design (subtest)

PUBLISHER OR AUTHOR: Psychological Corporation
ADDRESS: 304 East 45th Street
New York, New York 10017

CONTENT: Cognitive, Psychomotor

DESCRIPTION: Task consists of child's reproducing (constructing) designs with flat colored blocks either from examiner's model or picture on a card. The measure taps problem-solving abilities, flexibility of response style, visual-motor organization and execution.

AGE RANGE: 4 years through 6-1/2 years

ADMINISTRATION: Test time: 10 minutes (timed items)
Group or individual: Individual
Who administers: Trained tester
How administered: Testing
Response mode: Manual performance, manipulation of objects

VALIDITY: Studies have not fully examined the validity of this subtest.
Predictive: Studies with adequate sample sizes haven't been reported.
Concurrent: Correlations with the Stanford-Binet and the Full Scale I.Q. are reported in the .70-.80 range.

RELIABILITY: Test-retest reliabilities for Block Design for age 5 1/2 children is reported in the manual (r=.77). Block correlates with Full Scale I.Q. at .50-.65 at various ages.

NORMS: Norms were reported for representative selection of 1,200 U.S. children stratified by geographic region, ethnic background, and father's occupation.

SOURCE IN WHICH DESCRIBED: Wechsler Preschool and Primary Scale of Intelligence Manual (Wechsler, D. 1967)

PERMISSION TO USE CAN BE OBTAINED FROM: Psychological Corporation

Bilingual Syntax Measure

PUBLISHER OR AUTHOR: Marina Burt, Heidi Dulay and Eduardo Hernandez
ADDRESS: Harcourt, Brace, Jovanovich, Inc.
New York, New York

CONTENT: Language

DESCRIPTION: This test is designed to measure children's oral proficiency in English and/or Spanish grammatical structures by using natural speech as a basis for making judgments. Simple questions are used with cartoon-type colored pictures to provide the framework for a conversation with the child. An analysis of the child's speech yields a numerical indicator and a qualitative description of the child's structural language proficiency. Responses are written down verbatim.

AGE RANGE: 4 years through 9 years

ADMINISTRATION:
Test time: 10-15 minutes
Group or individual: Individual
Who administers: No training necessary (must be bilingual)
How administered: Interview
Response mode: Oral
Comments: This test can be used as a test of language dominance as well as a test of the degree of maintenance or loss of certain basic Spanish grammatical structures in children who acquired Spanish.

VALIDITY: High face validity. The Rationale and Technical Report was not available at the time of this writing.

NORMS: Norms on 2,300 English-speaking children and 2,000 Spanish-speaking children tested in March, 1974 are available from the publishers.

PERMISSION TO USE CAN BE OBTAINED FROM: Harcourt, Brace, Jovanovich

CIRCUS, subtest: How Much and How Many

PUBLISHER OR AUTHOR: Educational Testing Service
ADDRESS: Princeton, New Jersey

CONTENT: Cognitive

DESCRIPTION: Child is presented with sets of pictures and must choose the picture in each set which illustrates the appropriate numerical or relational concept. Authors claim the test measures these concepts and child's comprehension of the vocabulary used in the items.

AGE RANGE: Preschool and kindergarten

ADMINISTRATION: Test time: 20 minutes
Group or individual: Group
Who administers: Teacher
How administered: Testing
Response mode: Pencil and paper
Comments: Quantative concepts measure is confounded with vocabulary comprehension.

VALIDITY: No studies reported. Validity claimed by authors on the importance of the variables measured (face validity).

RELIABILITY: Reported split-half: r=.88, .87
Reported Alpha: r=.87, .88

NORMS: Standardization on 567 kindergarten children and 582 nursery school children from four geographic areas of the U.S. Since actual sample was based on returns from selected classes, sample is biased by those tests which were returned (i.e., White, middle-class).

SOURCE IN WHICH DESCRIBED: CIRCUS Manual and Technical Report (Anderson, et al., 1974)

PERMISSION TO USE CAN BE OBTAINED FROM: Educational Testing Service

CIRCUS, subtest: Finding Letters and Numbers

PUBLISHER OR AUTHOR: Educational Testing Service
ADDRESS: Princeton, New Jersey

CONTENT: Cognitive

DESCRIPTION: Child is presented with sets of pictures of letters or numbers and chooses the appropriate response representing the letter or number the teacher has named. This test is designed to measure recognition of letters and numbers, and as such is a readiness test.

AGE RANGE: Preschool and Kindergarten

ADMINISTRATION: Test time: 20 minutes
Group or individual: Group
Who administers: Teacher
How administered: Testing
Response mode: Child points or names correct item
Comments: Possibly useful as a diagnostic test, although sequencing, an important correlate of recognition of letters and numbers, is not tested.

VALIDITY: Not available (see previous page).

RELIABILITY: Reported split-half: r=.86

NORMS: Standardization on 290 kindergarten children and 546 nursery school children from four geographic areas of the U.S. Since actual sample was based on returns from selected classes, sample is biased by those tests which were returned (i.e., White, middle-class).

SOURCE IN WHICH DESCRIBED: CIRCUS Manual and Technical Report (Anderson, et al., 1974)

PERMISSION TO USE CAN BE OBTAINED FROM: Educational Testing Service

CIRCUS, subtest: How Words Work

PUBLISHER OR AUTHOR: Educational Testing Service
ADDRESS: Princeton, New Jersey

CONTENT: Cognitive

DESCRIPTION: Child listens to teacher reading a sentence and then chooses the appropriate picture which illustrates the functional words in that sentence. Functional words include prepositions, conjunctions, and pronouns. Authors describe this task as tapping a number of aspects of functional language.

AGE RANGE: Preschool and kindergarten

ADMINISTRATION: Test time: 20 minutes
Group or individual: Group
Who administers: Teacher
How administered: Testing
Response mode: Pencil and paper
Comments: Multiple concepts in each item and necessity for vocabular comprehension make actual scores impossible to interpret.

VALIDITY: Not available

RELIABILITY: Reported split-half: r=.81
Reported Alpha (internal consistency): r=.78

NORMS: Standardization on 252 kindergarten children and 594 nursery school children from four geographic areas of the U.S. Since actual sample was based on returns from selected classes, sample is biased by those tests which were returned (i.e., White, middle-class).

SOURCE IN WHICH DESCRIBED: CIRCUS Manual and Technical Report (Anderson, et al., 1974)

PERMISSION TO USE CAN BE OBTAINED FROM: Educational Testing Service

CIRCUS, subtest: Do You Know..?

PUBLISHER OR AUTHOR: Educational Testing Service
ADDRESS: Princeton, New Jersey

CONTENT: Cognitive

DESCRIPTION: This is a general information test. Child chooses appropriate picture which answers teacher's question. This task taps the child's experience in a variety of areas (health, safety, social standards, consumer concepts).

AGE RANGE: Preschool and kindergarten

ADMINISTRATION: Test time: 20 minutes
Group or individual: Group
Who administers: Teacher
How administered: Testing
Response mode: Paper and pencil
Comments: Obvious ceiling on a majority of items for the standardization sample. No examination of possible cultural bias in questions and pictures.

VALIDITY: Not available

RELIABILITY: Reported split-half: r=.78
Reported Alpha (interal consistency): r=.79

NORMS: Standardization on 286 kindergarten children and 591 nursery school children from four geographic areas of the U.S. Since actual sample was based on returns from selected classes, sample is biased by those tests which were returned (i.e., White, middle-class).

SOURCE IN WHICH DESCRIBED: CIRCUS Manual and Technical Report (Anderson, et al., 1974)

PERMISSION TO USE CAN BE OBTAINED FROM: Educational Testing Service

CIRCUS, subtest: Say and Tell

PUBLISHER OR AUTHOR: Educational Testing Service
ADDRESS: Princeton, New Jersey

CONTENT: Cognitive

DESCRIPTION: Part I of this test consists of two parts and taps children's descriptive language abilities. In the first part the child is given a pencil and asked attribute questions, e.g., "What color is it?"; in the second part the child is given two pennies and is asked to describe them. Scoring is based on categories of attribute which the child mentions.

AGE RANGE: Preschool and kindergarten

ADMINISTRATION: Test time: 5 minutes
Group or individual: Individual
Who administers: Teacher
How administered: Testing
Response mode: Oral
Comments: Reported test scores on the norming sample of nursery school and kindergarten children fail to show improved performance with age.

VALIDITY: Not available

RELIABILITY: Reported split-half: Not available
Reported Alpha (internal consistency): .72

NORMS: Standardization on 541 kindergarten and 227 nursery school children from four geographic areas of the U.S. Since actual sample was based on returns from selected classes, sample is biased by those tests which were returned (i.e., White, middle-class).

SOURCE IN WHICH DESCRIBED: CIRCUS Manual and Technical Report (Anderson, et al., 1974)

PERMISSION TO USE CAN BE OBTAINED FROM: Educational Testing Service

Auditory Discrimination Test (ADT)

PUBLISHER OR AUTHOR: Language Research Association
ADDRESS: 950 East 59th Street, Box 95
Chicago, IL 60637

CONTENT: Psychomotor

DESCRIPTION: Pairs of words, some the same, some different, are read to the child who is asked to respond "same" or "different" to each pair. A separate score is obtained for errors on both types of pairs.

AGE RANGE: 5 to 8 years

ADMINISTRATION: Test time: 5 minutes
Group or individual: Individual
Who administers: Teacher
How administered: Testing
Response mode: Oral

VALIDITY: Validity is determined by performance on same word pairs; auditory discrimination is determined by performance on different word pairs. Correlation with intelligence tests is reported as r = .32.

RELIABILITY: Test-retest reliability on 109 subjects: r = .91.

NORMS: The test has been normed on 533 children aged 5 years and older from both urban and non-urban backgrounds.

SOURCE IN WHICH DESCRIBED: Buros Sixth Mental Measurements Yearbook

PERMISSION TO USE CAN BE OBTAIN FROM: Language Research Association
950 East 59th Street, Box 95
Chicago, IL 60637